INTERNATIONAL FOOD
POLICY RESEARCH INSTITUTE
sustainable solutions for ending hunger and poverty
Supported by the CGIAR

2020 VISION

IFPRI Discussion Paper 00929

November 2009

Improving the Proof

Evolution of and Emerging Trends in Impact Assessment Methods and Approaches in Agricultural Development

Mywish K. Maredia

2020 Vision Initiative

This paper has been prepared for the project on
Millions Fed: Proven Successes in Agricultural Development
(www.ifpri.org/millionsfed)

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI) was established in 1975. IFPRI is one of 15 agricultural research centers that receive principal funding from governments, private foundations, and international and regional organizations, most of which are members of the Consultative Group on International Agricultural Research (CGIAR).

FINANCIAL CONTRIBUTORS AND PARTNERS

IFPRI's research, capacity strengthening, and communications work is made possible by its financial contributors and partners. IFPRI receives its principal funding from governments, private foundations, and international and regional organizations, most of which are members of the Consultative Group on International Agricultural Research (CGIAR). IFPRI gratefully acknowledges the generous unrestricted funding from Australia, Canada, China, Finland, France, Germany, India, Ireland, Italy, Japan, Netherlands, Norway, South Africa, Sweden, Switzerland, United Kingdom, United States, and World Bank.

MILLIONS FED

"Millions Fed: Proven Successes in Agricultural Development" is a project led by IFPRI and its 2020 Vision Initiative to identify interventions in agricultural development that have substantially reduced hunger and poverty; to document evidence about where, when, and why these interventions succeeded; to learn about the key drivers and factors underlying success; and to share lessons to help inform better policy and investment decisions in the future.

A total of 20 case studies are included in this project, each one based on a synthesis of the peer-reviewed literature, along with other relevant knowledge, that documents an intervention's impact on hunger and malnutrition and the pathways to food security. All these studies were in turn peer reviewed by both the Millions Fed project and IFPRI's independent Publications Review Committee.

AUTHORS

Mywish K. Maredia, Michigan State University
Associate Professor, Department of Agricultural, Food and Resource Economics
Email: maredia@msu.edu

Notices

[1] Effective January 2007, the Discussion Paper series within each division and the Director General's Office of IFPRI were merged into one IFPRI–wide Discussion Paper series. The new series begins with number 00689, reflecting the prior publication of 688 discussion papers within the dispersed series. The earlier series are available on IFPRI's website at www.ifpri.org/pubs/otherpubs.htm#dp.

Contents

List of Tables

List of Figures

List of Boxes

ACKNOWLEDGEMENTS

Comments and feedback received from the editors and anonymous reviewers in revising this paper are greatly appreciated. All the normal disclaimers apply.

ABSTRACT

Assessing impacts of public investments has long captured the interest and attention of the development community. This paper presents the evolution of different methods and approaches used for *ex ante* appraisal, monitoring, project evaluation, and impact assessment over the last five decades. Among these tools, impact assessment (IA) conducted retrospectively comes closest to providing the proof of development effectiveness. It is defined as the systematic analysis of the significant or lasting changes in people's lives brought about by a given action or series of actions in relation to a counterfactual. There are three basic types of retrospective IAs: macro-level IAs that focus on the contribution of developmental efforts to an impact goal aggregated at a sector or a system level; micro-level impact evaluations (IEs) concerned with estimating the average effect of an intervention on outcomes at the beneficiary level; and micro-level *ex post* impact analysis concerned with total effects of a development effort after the outputs are scaled-up. *Ex post* IAs have evolved and expanded over the decades in both breadth and depth of analysis in response to evolving development themes and methodological advancements. The increased emphasis on learning from evaluations has also seen responses from both quantitative and qualitative camps of the evaluation community. The paper argues that generation of robust knowledge that feeds into making developmental policies and investment decisions requires a hierarchical and cumulative approach to "improving the proof" through rigorous and a variety of impact assessment methods applied incrementally at the project, program and system level. Subjecting as many development interventions as resources allow to rigorous impact assessment based on a common framework can help build a critical body of evidence on impacts of development interventions, which can then be subjected to meta-analyses to help assimilate results across different studies and build a knowledge base on what works and what does not.

Keywords: Millions Fed, Food Security, Impact Assessment, Agriculture, Agricultural Development, Proof, Impact Evaluation

1. INTRODUCTION

The goal of any development effort is to achieve positive impacts on people's lives. Throughout the past several decades, this has served as a key objective in guiding the planning and implementation of a broad range of global and national development efforts. In the area of agriculture, major attention focuses on achieving people-related impacts, such as reducing poverty, hunger and food insecurity.

Whether, what and how development efforts have an impact on people's lives have long been topics of research and enquiry by academicians, development practitioners, investors, and target beneficiaries of development efforts. People have formally or informally, systematically or haphazardly, scientifically or unscientifically assessed, analyzed, measured, estimated, and evaluated impacts for as long as there have been development efforts. The motivation to assess impacts stems from the need for accountability (whether and what impacts development efforts have on people and their environment) and interest in institutional learning (how impacts are achieved or not achieved and what lessons can be derived to improve programs). Assessment and documentation of impacts provide the "proof" that development does or does not work and in what contexts.

With the establishment of the Millennium Development Goals (MDGs) and the promotion of the concept of "evidence-based policy" (stemming from the Paris Declaration), accountability and applying lessons learned from interventions have received an enormous boost in recent years. The need to be accountable for results and an increasingly urgent search for evidence on what works and what does not have increased the interest and attention on impacts. It has led to a growing focus and emphasis on a more formal, systematic, and rigorous approach to assessing, evaluating, estimating, and analyzing impacts of development projects, including agriculture and rural development (Goldstein et al. 2008, IFAD 2003). A result of this growing focus on assessing impacts is an interest in the evolution of the field of impact assessment itself. The motivation for this paper stems from this interest to better understand the theory and practice of impact assessment, especially, as it relates to agricultural development. It also attempts to clarify and define different concepts/terms commonly used in the impact assessment literature to better understand their contributions towards "improving the proof." In this context it is worth noting that the term "impact assessment" as used in this paper refers to the umbrella concept encompassing all analyses focused on outcomes and impacts.

The goal of this paper is to present an overview of how methods/approaches to impact assessment of agricultural development efforts have evolved over the past few decades, discuss the current practices and state of the art, and highlight new and emerging trends. Specific objectives are:

1. To examine the evolution of the methods and approaches found in the impact assessment literature of agricultural development in general, and as it relates to the goals of enhancing food security, in particular; and
2. To derive lessons for best practices and implications for enhancing the culture of impact assessment in agricultural development.

Toward these objectives, the paper first presents the conceptual framework that links agricultural development efforts with the impact goal of enhancing food security and puts it in the context of broader developmental goals. It also clarifies definitions of key concepts and draws a nuanced distinction between terminologies commonly used in the field of impact assessment of development programs, such as "impact evaluation" and "*ex post* impact assessment." This is followed by a historical overview of the evolution of research on impacts in light of the shifts in the development paradigm over the last six decades, the emerging trends in impact assessment methods and approaches, and their pros and cons in improving the proof. The review then focuses on past evidence, future potential and best practices in the assessment of developmental impacts related to food security. The conclusion section discusses lessons learned from this exercise and draws implications for improving the culture and strategy of impact assessment in agricultural development.

2. SETTING THE STAGE: DEFINING THE CONCEPTS

According to the latest figures available, the total annual donor commitment to agricultural development in 1990s and early 2000s was in the range of $10–12 billion (FAO 2005).[1] Although a breakdown of these investments by program goals is not available, it can be safely assumed that a majority of these development efforts in agriculture are devoted to achieving goals related to ending poverty and hunger. Enhancing food security is intrinsically linked with the achievement of these developmental goals and is thus considered one of the major focuses of agricultural development interventions. For the purposes of this paper, development interventions in the category of agriculture are broadly defined to include rural development, infrastructure development that affects agriculture directly (roads, irrigation, drainage systems), policy changes (as related to rural land, labor, capital, outputs, inputs, and prices), agricultural services in rural areas encompassing marketing and financial systems, research and provision of improved technologies, practices and other inputs (fertilizer, water, pesticides, seeds), and the provision of extension services to producers, among others.

Since "food security" and "impact assessment" are main themes of this paper, it is necessary to look closely at these concepts to establish the explicit or implied definitions.

Food Security

Several hundred definitions of food security are found in the literature as a result of the evolution in thinking and understanding by the international community of the complexities involved in the technical and policy aspects related to this concept. The official definition of food security has evolved from the initial focus in the 1970s on food supply problems – of assuring availability and to some degree price stability – to the focus in the 1980s and 1990s shifting toward the demand side and issues of consumption and access for vulnerable people at all times (see FAO 2003 for this historical perspective on the evolution of the concept of food security). In this paper we view food security as a phenomenon relating to individuals. Following FAO (2003), useful working definitions of food security and food insecurity implied in this paper are:

- *Food security* exists when all people, at all times, have physical, social, and economic access to sufficient, safe, and nutritious food which meets their dietary needs and food preferences for an active and healthy life. Household food security is the application of this concept to the family level, with individuals within households as the focus of concern.
- *Food insecurity* exists when people do not have adequate physical, social, or economic access to food as defined above.

Based on these concepts, the ultimate goal of agricultural development focused on increasing food security (or reducing food insecurity) is the enhancement of nutritional status of the individual household member, and reducing the risk of that adequate status not being achieved or becoming undermined. Over the last four decades, the practical response from the international community toward achieving this concept of food security has been to focus on simpler objectives around which to organize international and national public action, such as reduction and elimination of poverty. According to FAO (2003), the United Nations exemplified this direction of policy for international action on food security by setting the targets for the first MDG as halving the number of hungry and poor people by 2015. In this paper, therefore, enhancing food security is interpreted as implicitly embedded in the development goals of reducing poverty and hunger. In other words, development efforts targeted to reducing hunger and poverty are viewed as contributing towards the goal of enhancing food security. The overview of impact

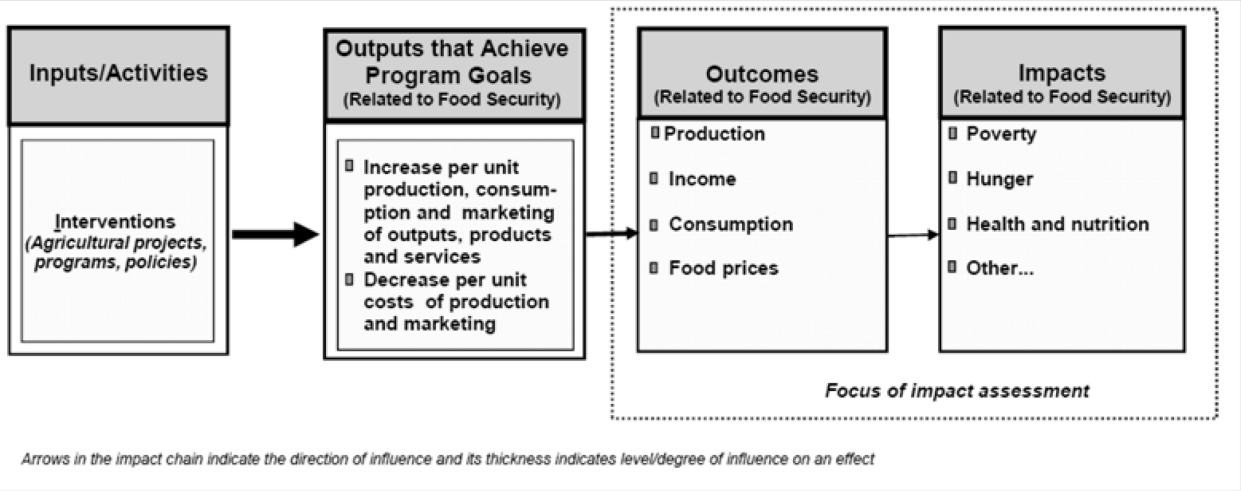
[1] This figure is an underestimate of the total global effort on agricultural development as it does not include developing countries' own support and commitment to agriculture.

assessments in agricultural development efforts presented in this paper takes this broader perspective of the goal of enhancing food security.

Impact Assessment

Impact assessment (IA), as used in this paper, is the systematic analysis of the significant or lasting changes—positive or negative, intended or not—in people's lives brought about by a given action or series of actions in relation to a counterfactual (adapted from Roche 1999). Figure 1 illustrates a simplified and generalized impact pathway (or a results chain) of how actions related to agricultural development (also referred as an intervention in the form of a program, project, policy change or an activity) affect the goal of enhancing food security. It also introduces the concept of "impact assessment," which is concerned with the evaluation of the final effects (long-term impacts on poverty, hunger, etc.) and intermediate effects (medium-term outcomes on production, income, consumption, and prices) *caused* by a given activity (Baker 2000).

Figure 1. A generalized impact pathway of agricultural interventions focused on enhancing food security



The impact pathway (or results chain) provides a useful way of conceptualizing the cause-and-effect relationship between different types of changes, with impact assessment focusing mainly on the changes occurring at the outcome and impact level. However, there are several concepts and terminologies found in the literature that are closely associated with impact assessment or that belong to the family of "Monitoring and Evaluation (M&E) and Impact Assessment." These are introduced in Annex 1, which combines the concept of impact chain with the timeline of project cycle to provide the classical view of the differences between appraisal, monitoring, and evaluation, and illustrates where impact assessment fits in this two-dimensional view of the world.

According to this classical view of the project cycle, monitoring, project evaluation and impact assessments are shown to occur in a specific time frame and focus on specific nodes of the impact chain. However, in practice the time frame of these assessments and the impact chain focal points may not be so well-defined. Although retrospective impact assessment occurs ex post of an intervention, the planning of such IA as an activity needs to occur concurrently with project planning and implementation so that baseline data are collected and to ensure indicators have been monitored efficiently. Moreover, although IA focuses mainly on the outcome/impact nodes of the causal chain, it needs to explore the entire chain if reliable conclusions can be drawn about the degree to which any observed change can be attributed to a given intervention. This means that impact assessment as an activity must be done throughout the project cycle; what changes is the nature of the exercise (Roche 1999). In the appraisal stages, it is anticipatory;

in the implementation stages, it checks what consequences a project is currently having; and after completion, the emphasis is on examining what effects the project has had.[2] A feedback loop connects all types of assessments (*ex ante*, monitoring, project/program evaluation, and *ex post* IA) with each other, and although the retrospective IA comes closest to providing the proof, it is the close integration and linkages between the assessments occurring at different stages of a project cycle that improves project effectiveness.

Thus, although retrospective impact assessments concerned with providing a proof of development effectiveness are the major focus of this paper (the box highlighted in yellow in Annex 1, Figure A1), *ex ante* impact assessments and other types of monitoring and evaluation assessments done throughout the continuum of impact pathway and project life cycles are included in the discussion where appropriate. The historical overview and the discussion of methods/approaches presented in the following sections introduce these different concepts/methods related to impact assessment with the hope of not only clarifying their meanings but also understanding their contributions toward "improving the proof."

Impact Assessment of Food Security Goal

Conceptually, the interventions (agricultural programs, projects, policies, and activities), as the first node in the results chain, attempt to bring about changes in the use of farm- and community-level resources and assets (land, labor, capital, entrepreneurship) to increase per unit production or marketing of outputs, products and services, or decrease per unit costs at the farm household level (referred in Figure 1 as program goals resulting from project outputs). These affect four major indicators that are directly linked with the goal of enhancing food security: production, income, consumption, and food prices.

In theory, increase in income and consumption and reduction in food prices serve as impact channels on the demand side (they increase the effective demand and intake of food) and increase in agricultural production serves as an impact channel that affects the supply of food. However, both in theory and practice, food production affects the supply as well as the demand side indicators related to income, consumption, and food prices. For example, an increase in production can increase producers' income, which can increase consumption. However, an increase in production can also reduce price and thus farmer income; but it can increase the real wages in the economy and positively affect consumption of net food buyers. Sorting out the net impact of any outcome on the goal of food security is thus a complex issue.

Another source of complexity is that agricultural development efforts can *effect* changes (positive or negative, intended or not) [3] in the ultimate impact goals of poverty, hunger, health, and nutrition at individual level, community level, or an aggregate national level. One of the limitations of the simple two-dimensional model of the impact pathway depicted in Figure 1 is that it does not include the size and scale dimension for either the development effort or that of the outcome and impact indicators. From the perspective of a single intervention, the channel of impact transmission can be viewed as influencing impact indicators at the micro-level (individuals, households, or a community). However, taking the perspective of a development goal at an aggregate level (reflected in macro-level indicators), the generic pathway can be viewed as resulting from collective changes in impact indicators aggregated across a large number of individuals across communities, over a sustainable period of time, and resulting from a multitude of interventions. The tapering thickness of the arrows along the results chain in Figure 1 indicates this reducing influence of a single intervention on aggregate level outcomes and impacts.

The conceptual framework depicted in Figure 1 also does not reflect the stage when impact

[2] In practice this is possible only for IAs that are project-specific. As the level of aggregation of an impact assessment increases from a project to a program to a system level, it becomes less practical to plan and conduct IAs concurrently with planning and implementation phases.

[3] The intended positive effects for the beneficiaries are the goals of the intervention. The unintended are positive or negative side effects. These welfare effects can be directly caused by the intervention or indirectly linked to the intervention through changes in the environment or the resource base for societies, institutions, groups or individuals. All these types of effects fall within the scope of impact assessment.

assessment is conducted, which could be at the initial pilot or "proof of concept" stage, or at the stage when project outputs are already scaled up and scaled out. From which perspective an impact assessment is conducted (from a single project perspective or a development goal perspective) and the stage in the impact pathway when it is conducted (initial pilot stage or after the program/policy is scaled up) have implications on the methods and approaches used for providing the proof on development effectiveness. The historical overview presented in this paper clarifies and distinguishes among the following broad categories of impact assessments that occur ex post of an activity, and which contribute to providing different types of proofs of development effectiveness.

Defining Micro- and Macro-Level IA, Impact Evaluation and *Ex Post* Impact Analysis

First, two broad categories of impact assessments (relevant to agricultural development) are distinguished in this paper for discussion purposes:[4]

a) ***Macro-level IA****:* IAs that are macro-level aggregate studies focus on assessing the contribution of past investments in a specific category of agricultural development (infrastructure, research, extension, etc.) or a sub-sector (crop, livestock, agro-forestry) to realized macro-level goals. These types of assessments provide evidence in relation to the long-term effectiveness of broad sector-level investments. They cannot provide information on the contribution of or effectiveness of specific project/program/intervention to a developmental goal.

b) ***Micro-level IA:*** IAs that are micro-level are intervention-specific and trace the inputs-outputs-outcomes-impacts relationship along the impact pathway from left to right for a specific activity or a group of activities. Micro-level IAs establish the causal link between inputs-outputs-outcomes-impacts in order to attribute the estimated/observed impacts to a specific intervention. Thus, they are best suited to providing proof of effectiveness of a specific intervention in achieving medium-term outcomes and development goals at the beneficiary level or a long-term goal aggregated across groups of beneficiaries.

Micro-level impact assessments typically involve four sequential steps.[5] Some of these steps may be more relevant for some IAs than others (as noted in the parenthesis below), depending on the stage of the intervention being assessed (either the pilot or "proof of concept" stage, or the stage when project outputs are already scaled up) and the motivation or questions being addressed by an IA (for example, how much better off or worse off are beneficiaries because of the intervention versus what are the total program benefits compared to program costs).[6]

Step 1: (All) Select the intervention (a project, program, policy change, activity) to be assessed and identify its outputs.

Step 2: (All) Estimate the "average size of the effect"[7] of the output at the level of individuals, households, or communities affected by the project output.[8]

[4] These correspond to the distinction between "aggregate" and "disaggregate" categories of ex post impact assessment as it relates to agricultural research interventions. Based on the primary objective of the IA, Walker et al. (2008) further group these two in to economic rate of return (IAs focused on productivity and profitability) and multi-dimensional impact studies (IAs focused on higher order impacts such as poverty, heath, environment).

[5] See Maredia (2009) for the application of these steps to agricultural research for development (R4D) interventions.

[6] Typically, all these steps have a temporal dimension which is not included in this discussion for the sake of simplicity.

[7] This is one of the most important concepts related to impact assessment. If, for example, the outcome indicator of interest is income, then the average effect is defined as the difference between the expected income earned by project beneficiaries while participating in the project (factual scenario) and the expected income they would have received if they had not participated in the project (counterfactual scenario).

[8] Note that depending on the source of data used to estimate the average size of the effect, this step could precede or follow step 3, or could occur in conjunction with step 3.

Step 3: (For assessments that occur at a stage when project outputs are already scaled up) Identify the domain of affected area and estimate its size/scale in units of geographic area, households, or people.

Step 4: (For assessments that occur at a stage when project outputs are already scaled up) Estimate the total size of the effect of an intervention (the benefit) as a function of average effect size (E_s, estimated in step 3) and effect scale (E_c, estimated in step 2). In other words, estimate project benefits, B, as: $B = f(E_s * E_c)$

Step 5: (Optional) Assess the impact on macro-level developmental goals as a function of results of step 4 (and other variables).

Step 6: (Optional) Compare the benefits estimated from steps 4 or 5 with the costs of realizing these benefits. This step is not always needed, but is useful to demonstrate the cost-effectiveness of an investment decision.

In the context of these steps, the following two types of micro-level IAs are differentiated in this paper based on the focal steps and motivation of the impact assessment:

b1) ***Impact evaluation (IE):*** are studies that focus on steps 1-2 and are mainly concerned with assessing the efficacy of an intervention. These evaluations address the question, What is the effect of an intervention on outcome indicators at a typical beneficiary level?.

b2) ***Ex post impact analysis (epIA):*** are concerned with documenting the evidence of total impacts realized across all beneficiaries (steps 1-4) and are driven by the demand for evidence of net benefits or returns to total investment in an activity.

Impact evaluations are concerned with estimating the "average effect" (E_s) of an intervention, and they are often needed as a proof of development efficacy prior to scaling up the project or when it is implemented at a small scale (Duflo et al. 2008, Ashraf et al. 2008). On the other hand, *ex post* impact analyses (also referred as *ex post* impact assessment in the agricultural research evaluation literature and a term used in this paper to distinguish it from IEs)[9] are concerned with estimating the "total effect" ($E_s * E_c$) after a project (or its output) is scaled up and scaled-out (as exemplified, for example by Rutherford 2008; Zeddies et al. 2001). These latter types of IAs are more concerned with providing proof of total benefits to justify past investments with the hope of new or continued investments in specific types of development interventions.

A key concept underlying the estimation of E_s (average effect size) is the counterfactual outcome (the predicted outcome in the absence of the intervention, or what would have happened to the beneficiaries had they not participated in the project). Unfortunately, the counterfactual outcome is never actually observed as people cannot simultaneously participate and not participate in a program. The credibility of how the counterfactual outcome is identified and estimated is at the heart of quantitative evaluation designs. The methods and approaches underlying the different types of impact assessments (macro-, micro-, IE and epIA) and the types of evaluation designs concerned with establishing the counterfactual are discussed in the following section in the context of the evolution of impact assessment methods and approaches and further elaborated in the context of the assessment of food security impacts.

[9] In this context the term "impact evaluation" as used in this paper refers to this specific type of impact assessment.

3. IMPROVING THE PROOF THROUGH IMPACT ASSESSMENT: THE EVOLUTION OF DIFFERENT METHODS AND APPROACHES

The theory and practice of appraisal, evaluation, and impact assessment in agricultural development is very much linked with the theory and practice of these assessments in the broad field of development (that includes other sectors such as education, health, environment).[10] The historical overviews of IA provided by Roche (1999) and Howes (1992) serve as major sources of background material for this discussion. Their overviews are extended in this paper to include the experience of past two decades and supplemented with examples from agricultural development. The historical overview of impact assessment methods and approaches presented below is cast in the light of the evolution of major development paradigms, themes, and ideas in agricultural and rural development over the past half-century.

Table A2 in Annex 2 summarizes the evolving and dominant theories, ideas, and policies that have guided rural development efforts over the past six decades. It shows the evolution of the dominant ideas of modernization, dual economy, and community development in the 1950s and 1960s that led to the top-down or "blueprint" approach to rural development, characterized by external technologies and national level policies, to the bottom-up, grassroots, or "process" approach in the late 1970s to early 1990s (Ellis and Biggs, 2001). The latter approach viewed rural development as a participatory process that empowers rural dwellers to contribute to solutions to the problems they face. The past decade or so has seen the emergence of a sustainable livelihoods framework and mainstreaming rural development based on ideas of innovation systems, participation with results, evidence-based policymaking and poverty reduction strategy papers (see Table A2 in Annex 2).

This broad historical perspective on the evolution of major ideas and themes in rural development are applied to understand the trends and changes in the tools, methods, and approaches to impact assessment, which is grouped into two categories—retrospective IA, and all the other types of assessments that occur during project planning/appraisal and implementation stage (*ex ante* IA and M&E). Figure 2 illustrates the emerging trends and state of the art in impact assessment methods, and the evolution of these various approaches in relation to major development paradigms grouped as modernization, limited participation, extended participation, and sustainable livelihoods. Modernization in this context refers to an approach largely premised on promoting economic and infrastructural development as a means for developing countries to catch up with the industrialized world. On the other hand, the extended participation approach begins with the belief that overcoming poverty is impossible without people's full participation. This paradigm is based on the premise that outsiders relinquish control and act as catalysts for locally owned processes of empowerment and development. The limited participation approach, in Howes' (1992) view, represents a compromise between these two poles and was most apparent in the multilateral agencies' shift to embrace participatory approaches, while retaining a strong planning tradition and an emphasis on economic development. The sustainable livelihoods approach represents the more recent paradigm shift based on the view that agriculture takes its place along with a host of other rural and non-rural activities that are important to the construction of viable rural livelihoods. Cross-sectoral and multi-sectoral diversity of rural livelihoods has become the cornerstone of rural development policy to reduce and eradicate rural poverty.

Evolution of IA Methods and Approaches: The *Ex Ante* and M&E Perspective

The initial use of systematic tools of impact assessment dates from the 1950s when international development was dominated by multilateral and bilateral donors who were guided exclusively by

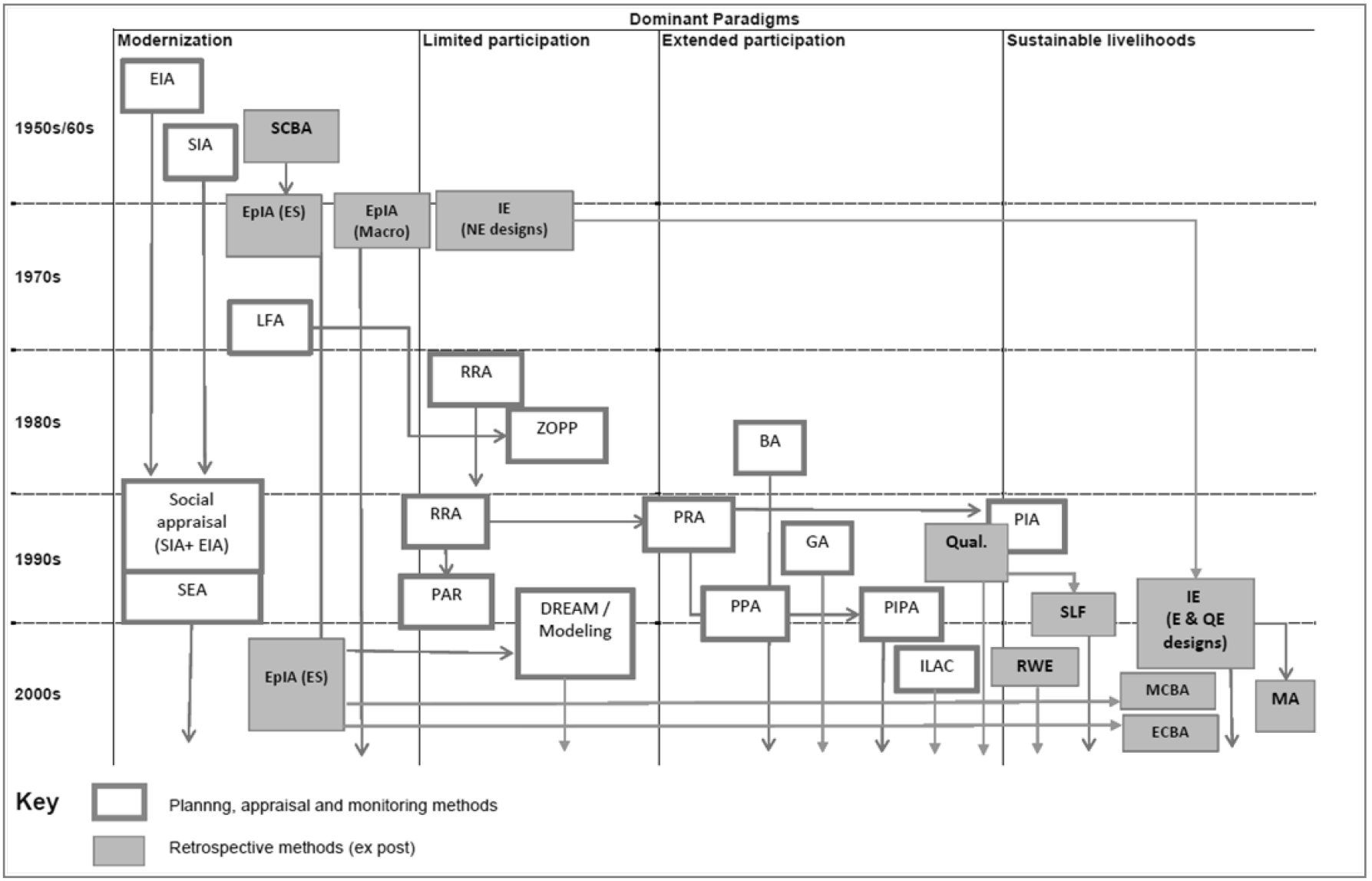
[10] This is not only because there are inter-sectoral linkages between these broad areas of development (and thus they share common principles and theory of development), but also because the same investors and donors fund development activities in these different areas. Thus there is a great potential for spillovers in the theory and practice of impact assessment across development interventions.

modernization theory. Projects were regarded as the primary means by which development was to be achieved. Development agencies began to use *ex ante* IA and appraisal approaches to predict a project's likely environmental, social, and economic consequences as a condition to approve, adjust or reject the project funding.

Environmental impact assessment (EIA), social impact assessment (SIA), and cost–benefit analysis (CBA) (or social cost–benefit analysis—SCBA) are some of the most common approaches that evolved over time in this family of methods focused on *ex ante* assessments of impacts (Figure 2). These assessments use program theory, bio-physical and socio-economic models, tools and techniques to predict and project the type, size, and scale of impacts of planned investments in a development activity. In the area of agriculture, these assessments are routine (and perhaps mandatory) in large investment projects related to infrastructure development such as water/irrigation management projects (dams, for instance), construction of rural roads, and drainage systems. Over the decades, these methods and approaches have evolved and combined into social appraisal and extended to programs and policies, which s referred as strategic environmental assessment (SEA) to denote its application to a higher, earlier and strategic tier of decisionmaking (Therivel et al. 1992; Therivel and Partidario 1996).

The next generation of monitoring and evaluation tools and approaches saw the introduction of logical framework analysis (LFA) in the planning and appraisal of development projects (Figure 2). This occurred in the context of rising criticism that investments in large-scale modernization projects such as infrastructure development were failing to benefit the poor. Satisfying "basic needs" and targeting "more aid to the poorest" were identified as central organizing principles for rural development. This was accompanied by an increasing pressure to engage with indigenous institutions and for a more active role for client groups. The LFA framework, which was a major innovation in donor practice in the 1970s (after it was pioneered by the U.S. Agency for International Development) went some way toward encompassing these new realities for greater accountability and client participation. It attempts to set out a clear hierarchy of inputs, activities, and objectives and relates these to assumptions about the external environment. The log frame matrix reflects the underlying theory of change of how inputs transfer to impacts. In this sense, LFA has clear linkages with retrospective IAs as the theory of change is based on prior evidence of cause-and effect relationships. The log frames also serve to outline the impact pathway for IAs to be conducted *ex post*. Today, variants of this logical framework are the most common planning framework used by bilateral and multilateral agencies, including those in the area of agricultural development. This class of planning tools maps out the cause-and-effect relationships based on pre-formed theories of change, and serves as the basis for monitoring the progress toward achieving project objectives.

Figure 2: Evolution of planning, monitoring, and impact assessment methods and approaches – 1950s-2000s



Source: Adapted from Howes (1992) and Roche (1999) to include evolution and application of methods in agricultural development since mid-1990s

In the late 1980s and 1990s, with the development paradigm shifting toward more participation-based development, the German development agency introduced the notion of participation into the logical framework approach with its development of objectives-oriented project planning (ZOPP) (GTZ 1988, 1998) (Figure 2).

As described in Roche (1999) and Howes (1992), the 1980s saw the blossoming of new methods such as rapid rural appraisal (RRA), participatory action research (PAR), participatory rural appraisal (PRA), and other types of participatory methods (Chambers 1992, 1994). These new methods of research and enquiry sought to make people and communities subjects and active participants, rather than mere objects of development. The application of these methods and approaches focused predominantly on agricultural and rural development programs. Participatory appraisal methods used a combination of mapping, diagramming, and semi-structured interviewing methods that allowed outsiders to build quick rapport with rural people. These methods also encouraged active participation among project beneficiaries and different interest groups in reviewing problems and determining priority areas for action. Later, with the emergence of the livelihoods framework and the promotion of the culture of institutional learning, this has led to the next generation of participatory methods in impact assessment, referred as Participatory Impact Assessment (PIA) and Participatory Impact Pathways Analysis (PIPA) (Chambers 2007, Catley et al. 2007, Douthwaite et al. 2007).

Qualitative approaches to the evaluation of development started emerging around 1980s that were built on sociological, historical and anthropological ideas. These methods viewed evaluation more as a negotiation of diverse opinions and perspectives. Focus groups were commonly used as a forum for interviewing a number of beneficiaries and for conducting institutional assessment. Such beneficiary assessments (BA) are meant to complement quantitative surveys and other traditional methods for data gathering with reliable and useful information on the socio-cultural context and perceptions of a client population. These qualitative approaches thus seek to understand the opinions of various interest groups and stakeholders, especially those whose views are not normally heard. Gender analysis (GA), although not strictly an evaluation method, was mainstreamed in the planning, appraisal, and evaluation of development projects. National-level planning and development strategies have also begun to include participatory poverty assessments (PPA), a type of BA, seeking to incorporate local perspectives and opinions by including participatory research methods (Chambers 2007).

Institutional learning and change (ILAC) is yet another broad category of new and emerging methods that incorporate participatory approaches to monitoring and evaluation across different institutions and sectors (Guijt et al. 2000, Guijt 2007, Watts et al. 2003). ILAC's concept of participatory M&E is based on the belief that learning is important for social change to happen and this (learning) occurs very much by doing and being, by reflecting critically on experience and actions in the world, and by testing the understanding that emanates from this through practice. This is also the underlying philosophy of the movement occurring beyond the evaluation community known as "learning for social change" (Deak et al. 2006; Taylor and Fransman 2004).

In the past 10 to 15 years, the rise in fast computing technology and tools such as GIS have also given rise to modeling approaches in project planning and priority setting. In the context of agricultural research and development (R&D), DREAM (**D**ynamic **R**esearch **E**valu**A**tion for **M**anagement) has evolved as one of the leading tools for *ex ante* impact assessment (Figure 2). DREAM combines the economic surplus (ES) principles used in epIA to simulate a range of market, technology adoption, research spillover, and trade policy scenarios to project economic benefits of a proposed R&D investment (Alston et al. 1998).

The appraisal and monitoring methods that have evolved over the past five decades and become part of development interventions do not directly address the issue of providing or improving the proof. However, their use and integration in the planning and implementation of development projects/programs contributes to: a) ensuring that development interventions do have potential positive effects (or avoid investments with net negative impacts); b) documenting outputs and ensuring that activities are achieving milestones and benchmarks toward the projected impact pathway; and c) generating knowledge and

information on the process of generating impacts. These play an important role in increasing the effectiveness of development interventions and the potential for achieving impacts.

Evolution of IA Methods and Approaches: The *Ex Post* Perspective

In contrast to the *ex ante* impact assessment, retrospective impact assessments, or assessment conducted after an intervention has ended (up to 10–15 years later), was a phenomenon that began in late 1950s–1960s and has evolved and expanded over the decades in both breadth and depth of analysis. These types of IAs come closest to providing the proof of development effectiveness. The discussion that follows is therefore geared toward how the different methods and approaches address the issue of assessing development effectiveness in terms of food security goals and the emerging trends in "improving the proof."

The Expanding Agenda of epIAs

In the area of agricultural interventions related to the adoption of new technologies (that is, interventions related to agricultural research and extension), the *ex post* assessments of impacts began with the pioneering work by Griliches (1958) that examined the social benefits of investing in the hybrid corn technology in the United States. Using a social cost–benefit framework based on economic surplus estimation methodology, this class of impact assessments flourished with the advent of the Green Revolution in the 1970s and 1980s. Over the past five decades, hundreds of studies have documented economic returns to investment in agricultural research (see for example summary overviews of this literature by Alston et al. 2000 and Evenson 2001). Many of these studies provide evidence of high rates of return to agricultural research investments (mostly measured in terms of productivity impact indicators), which support expanded public investments in agricultural research. Over time the economic surplus (ES) studies based on social cost-benefit analysis (CBA) has expanded to include non-Green Revolution technologies and non-economic impact themes. In terms of the food security implications, the evidence of high rates of return from these IAs provides the proof of the positive causal link between development interventions that increase productivity and gains in producer and consumer welfare as a result of increased income (for producers) and lower prices (for consumers).

The social cost–benefit framework underlying the retrospective impact assessments of the types described above is designed to estimate relative efficiency, expressed as a ratio between the total values of the inputs and the effects/impacts generated from those inputs. The estimates of average effect (parameter E_s in step 2 noted earlier) to estimate the total impact have traditionally been derived using a variety of sources and methods, some rigorous than others, referred in Table 3 as IEs using non-experimental (NE) methods. These include extrapolating/inferring the effects from controlled agricultural experiments, econometric simulation models based on secondary data to construct a "without intervention" scenario, reflexive studies that compare beneficiaries before and after an intervention, or simply comparing the participants with non-participant groups. The recent push for experimental and quasi-experimental methods in IE (discussed below) seeks to improve the quality of the control/comparison group for estimating E_s. These methods are considered more rigorous than NE methods traditionally used in IAs that estimate total effects (that is, epIA).

The total "social" impacts estimated using the social cost–benefit framework refers to the notion of the value of the project to the society as a whole after accounting for the potentially distorting effects of factors such as subsidies on prices.[11] In theory, the assessment of the true value of the project to the society must include intended (planned) and unintended (externalities, spillovers), short-term and longer-term costs and benefits in social, economic and environmental dimensions of people's lives. In practice, however, such comprehensive and multi-dimensional assessments of an intervention's impacts are rare due to lack of data, methodological limitations, and limited resources. Although it is uncommon for a

[11] The word "social" has nothing to do with the normal sense of relationships between groups and individuals.

single assessment to be all inclusive and comprehensive, research on impact assessment over the last five decades has advanced substantially, resulting in a more comprehensive portfolio of studies over time.

Pingali (2001) provides an insightful overview of the expanding agenda for research on retrospective impact assessments as practiced within the international agricultural research community. The study traces the evolution of this research on IA (conducted *ex post*) as a logical progression from the relatively narrow focus in the 1970s and 1980s on adoption and assessments of improved germplasm and crop management technologies, to formal rate of return and benefit distribution studies starting in the 1980s. The next major broadening of the agenda occurred in the 1980s with the coverage on spillover and inter-sectoral impacts. The activity expanded to include gender, health, and natural resource management impacts in the 1990s, and has expanded further to include policy research, poverty, social, and environmental impacts in the late 1990s and 2000s (Kelley et al. 2008).

The expanding agenda of impact assessment of agricultural research also reflects the progression toward assessing more types of research (such as crop germplasm, crop management, NRM, and policy) as well as different types of impacts (economic ROR, spillover, equity, health, gender, social, and environmental). Thus, the expanding agenda of micro-level *ex post* IAs to capture total effects reflects the expanding criteria (or yardstick) for measuring development effectiveness. The impact assessment portfolio has thus expanded from providing the proof of development effectiveness measured only in terms of economic rates of return to measuring development effectiveness in terms of other dimensions such as poverty, health, gender, and the environment. This broadening agenda of retrospective impact assessments is a result of not only methodological advancements but also the evolution of development paradigms and ideas that have shifted away from the one-dimensional (economic), top-down view of development to a multi-dimensional, participatory and people-focused development (Figure A1). Goals such as poverty alleviation, environmental sustainability, and gender equity have come to the forefront of the development agenda, and investors are increasingly demanding accountability around achievement of these goals from all types of interventions. The emergence of extended cost–benefit analysis (ECBA) in recent years is a direct response to this growing demand (Figure 2). Although applied mainly in the context of *ex ante* analysis, ECBA has the potential to broaden the scope of *ex post* impact assessment to include environmental and social costs and benefits (Bennett 2009). The ECBA when applied to an *ex post* setting has the potential for providing the proof of not only "how many millions are fed" and "what are the social benefits," but also "at what environmental or social costs?"

Macro-level IAs

Although the micro-level (or intervention-specific) and macro-level IAs (aggregate goal-level assessment) have evolved in tandem over the past five decades, the methods and approaches used for impact assessment are quite different (Figure 2). Whereas the evaluation methodology for micro-level analysis is based on assessment of causal attribution, the methodology for macro-level analysis of *ex post* impacts depends on estimates based on causal contribution. Macro-level IAs typically take a statistical approach that relates changes in macro-level indicators (total factor productivity, poverty level) to some aggregate level indicators of "inputs" (usually, investments in a type of agricultural development or as sub-sector). They are based on secondary data and require specialized skills in statistical methods, particularly time series data analyses. In contrast, micro-level studies that focus on one or more well defined interventions draw on multiple data sources for evaluating benefits, invest in primary data collection, and interact with project teams directly involved in generating the outputs being assessed. Alston et al. 1998 provide an overview of methodologies applied to both macro- and micro-level impact assessment of agricultural research.

Classic macro-level IA studies focusing on economic impacts of sector-wide investments include Griliches (1963, 1964) and Evenson and Kislev (1975). The macro-level impact analysis has evolved over the decades to include other dimensions of impacts closer to food security goals such as poverty and nutrition (Fan 2007; Fan et al. 2007; Evenson and Rosegrant 2003). Such studies are useful in showing the cause-and-effect relationship between development efforts aggregated across a sector, type of effort,

or spatial and temporal dimension and macro-level indicators of impact goals (such as number of people lifted out of poverty, number of children saved from malnutrition, etc.). For example, Evenson and Rosegrant (2003) look at the impacts of increased productivity as a result of crop germplasm improvement research on aggregate level welfare indicators related to poverty and food security. After taking into account the general equilibrium effects of increased production and reduced prices on demand and supply of food, they report that in the counterfactual scenario (the absence of improved crop varietal technologies), world food production would have been 4–5 percent lower in developing countries, and 13–15 million more children would have suffered from hunger and malnourishment. A major limitation of these types of methods is that the effects measured by such analysis cannot be tied back to any specific intervention to assess what works and what does not. However, these types of studies contribute to the formation of theory of change and to a better understanding of the relative importance of broad categories of development efforts, such as agriculture research, infrastructure, and education, in effecting macro-level impact goals.

Shifting Development Paradigms and the Rise of Qualitative Techniques:

The shifting development paradigms from none or limited participation in the 1960s and 1970s to extended participation in the 1980s and 1990s that led to the participatory methods of appraisal and monitoring of development projects also led to the use of qualitative techniques (Qual. tech.) in retrospective impact assessments. Such qualitative assessments are based on the principles of participation and interpretation of information documented throughout the process of evaluation (See Box 1 for a description of this method). Because measuring the counterfactual is at the core of impact analysis techniques, qualitative designs have generally been used in conjunction with other evaluation techniques. Recent developments of this technique have also seen its application in CBA which is traditionally considered a quantitative method (e.g., Rogers et al. 2009).

Box 1: Qualitative impact evaluation methods

Qualitative techniques are used for carrying out impact assessment with the intent to determine impact by the reliance on something other than the counterfactual (Mohr 1995, 1999). In one sense, these methods use non-experimental and non-statistical methods. Attribution/contribution is assessed using approaches such as reference to secondary data, program theory (logic models), theory of change and concept mapping. The focus is on understanding processes, behaviors, and conditions as they are perceived by the individuals or groups being studied (Valadez and Bamberger 1994). The qualitative approach uses relatively open-ended methods during design, collection of data, and analysis. Qualitative data can also be quantified. Among the methodologies used in qualitative impact assessments are the techniques developed for rapid rural assessment, which rely on participants' knowledge of the conditions surrounding the project or program being evaluated, or participatory evaluations in which stakeholders are involved in all stages of the evaluation—determining the objectives of the study, identifying and selecting indicators to be used, and participating in data collection and analysis.

A recent trend in impact assessment that has elements of participatory monitoring as well as qualitative techniques of impact assessment is the use of the sustainable livelihoods framework (SLF) as the underlying theory of change and a tool for evaluation (La Rovere et al. 2008, Mancini et al. 2007, Adato and Meinzen-Dick 2007) (Figure 2). The evaluation framework is based on guiding project beneficiaries in conceptualizing changes over time in their overall livelihood. These types of assessments are multi-dimensional, qualitative, and present results in terms of indicators that measure capabilities and assets of participants/beneficiaries of an intervention. The evolution of this method and approach is a direct consequence of the emerging development paradigm emphasizing the sustainability aspects of people's livelihoods defined in terms of different types of assets held by a household.

The benefits of qualitative assessments are that they are flexible, can be specifically tailored to the needs of the evaluation using open-ended approaches, can be carried out quickly using rapid

techniques, and can greatly enhance the findings of an impact assessment through providing a better understanding of stakeholders' perceptions and priorities and the conditions and processes which may have affected program impact. Thus, in the context of food security goals, they are good at asking the "why" and "how" questions related to intra-household patterns of food consumption, nutrition, and distribution. Among the main drawbacks are the subjectivity involved in data collection, the lack of a comparison group, and the lack of statistical robustness, given mainly small sample sizes, all of which make it difficult to generalize to a larger, representative population. The validity and reliability of qualitative data are highly dependent on the methodological skill, sensitivity, and training of the evaluator. If field staff are not sensitive to specific social and cultural norms and practices and to nonverbal messages, the data collected may be misinterpreted. And finally, without a comparison group, it is impossible to determine the counterfactual and thus causality of project impact. In other words, qualitative assessments are not good at measuring whether a project has impacts and the total impacts attributed to a project.

The Push for Learning from Evaluations: Responses from Quantitative Methods

The decade of 2000s has also seen an increased emphasis on learning from evaluations, which parallels the emergence of ILAC in monitoring and project evaluation discussed before. The qualitative method (including participatory techniques) noted above is one of the responses toward this increased interest in learning through process evaluation (rather than simply summative evaluation). However, there is also an increased emphasis on learning from quantitative methods as well. This is reflected in the conclusion of the 2006 report by the Center for Global Development (CGD) which argued that despite billions of dollars being spent on health, education, and social development programs in developing countries, there is relatively little knowledge about the net impact of most of these programs (Center for Global Development 2006). This lack of evidence of what works and what does not, according to the CGD report, implies that critical knowledge is withheld from policymaking, thus reducing the effectiveness of publicly funded development programs. The sentiments expressed in the CGD report are shared by many funders and practitioners of development and are the foundation of the movement toward "evidence-based policy" witnessed in the last decade. As a result, rigorous quantitative methods for estimating counterfactuals such as experimental (E) and quasi-experimental (QE) methods, and the practice of synthesizing results from past evaluations and assessments through systematic reviews are being promoted in the practice of development (Figure 2). These are elaborated below.

1. *Meta-analysis:* Systematic syntheses and reviews (for example, as undertaken by "What Works Clearing House" in the area of education, and by the Campbell Collaboration in the areas of education, crime and justice, and social welfare),[12] meta-cost-benefit analysis (MCBA) (e.g., Raitzer and Kelley 2008, Maredia and Raitzer 2006) and statistical meta-analysis (MA) (e.g., Alston et al. 2000, IEG 2009) seem to be the new emerging trends in impact assessment of agricultural development interventions (Figure 2). These types of meta-syntheses not only help take a stock of what is known in IA in a given focused area of enquiry, but also help systematically build a body of evidence and identify gaps in knowledge about what works in development (and what does not). By systematically combining studies, one attempts to overcome limits of size or scope in individual studies to obtain more reliable information about the impact of a "treatment." Thus, these types of analyses have the potential for not only providing the proof but also improving the proof about what works and what does not in development at the micro-level.

   Among the different methods of synthesizing documented evidence from past evaluations, statistical meta-analysis has become increasingly popular in recent decades. Despite the controversy about its validity (see for example, Bailar 1997 and Thompson and

[12] The protocols, mission, and activities of these initiatives can be found at: http://ies.ed.gov/ncee/wwc/ and http://www.campbellcollaboration.org/.

Pocock 1991 cited in Berman and Parker 2002), it has been applied with increasing frequency in the area of education, health, and psychology (Glass 1976) and to some extent in the field of economic research (Espey et al. 1994, Holger and Strobl 2001, Thiam et al. 2001). But its application to evaluations of agricultural and rural development interventions is a recent phenomenon (Figure 2). With the advent of the application of experimental and quasi-experimental IE methods, scope exists to expand the use of meta-analysis techniques to improving the proof. This is because such quantitative methods provide a common methodological framework to synthesize findings across IEs.

2. *E and QE methods for estimating program effects:* Since the pioneering work by Fisher in 1930s, agricultural research has maintained a long tradition of using natural experiments to identify and test the efficacy of potential research outputs as a precursor to releasing technological outputs (such as improved varieties, agronomic practices, recommendations regarding level and application of inputs) in the public domain (Fisher 1935). However, despite the fact that the use of mathematical statistics in the design of experiments is very much an agricultural heritage, the use of social experiments to evaluate the impacts of specific interventions that combine agricultural technologies, practices and policies in a real world setting is still uncommon. Given the fact that agricultural and rural development is now viewed as an important contributor to the MDGs, rigorous methods of impact assessment, especially those related to Step 2 noted on page 8, are being promoted in agricultural development as well. Thus the last few years have seen an increased influence and use of IE methods such as experimental (E) designs (randomization) and quasi-experimental (QE) designs using statistical techniques to establish a comparison group (Figure 2). Baker (2000) and Ravallion (2008) provide excellent overviews of some of these quantitative methods of IEs (including some NE techniques such as the use of instrumental variables) concerned with establishing a counterfactual in a development setting.

   As summarized in Annex 3, the three quantitative evaluation designs (E, QE and NE) differ in the identification of a treatment/participant group and control/comparison group (and thus the counterfactual), how and when the evaluation is planned, and other elements of evaluation design which determine its rigor. The earlier an evaluation is planned in the context of a development intervention, the greater the methodological flexibility in terms of the choice of using experimental (and quasi-experimental) designs to ensure methodological rigor and robustness. This also explains why E and QE methods (to estimate the average effect size) have not been traditionally used in *ex post* impact analyses, which are typically conducted many years after the intervention is completed.

   In general, experimental designs are considered the most robust of the IE methodologies. Impact evaluations based on social experimental research design should provide a high level of internal validity for the causal link between an intervention and the social outcomes by overcoming the problem of selection bias that is endemic to non-randomized evaluations (see Box 2). Due to this advantage, social experiment has been advocated as the main tool for studying development effectiveness by an influential group of academic economists (called the "randomistas") led by the MIT's Poverty Action Lab and promoters of evidence-based policy making (Banerjee 2007, Duflo 2006, Duflo et al. 2007).[13]

[13] See the 'Evidence-Based Policy' website (www.evidencebasedpolicy.org) for resources and tools for social experimental designs.

Box 2: The problem of selection bias and solving it through experimental designs

There are two types of selection biases that experimental designs overcome (in theory)—bias in the selection of individuals/groups receiving an intervention (normally referred simply as "selection bias") and the bias in the selection of interventions for assessment and its publication (Duflo et al. 2007). The first type of selection bias arises when individuals or groups receiving an intervention are self-selected or are assigned to the treatment group based on characteristics that may also affect their outcomes. This makes it difficult to disentangle the impact of the intervention from the factors that drove selection. Quasi-experimental designs (including, non-randomized controlled trials) and non-experimental designs can detect associations between an intervention and an outcome. But they cannot rule out the possibility that the association was caused by a third factor linked to both intervention and outcome. Random allocation ensures that there are no systematic differences between intervention groups in factors, known and unknown, that may affect outcome (for a mathematical treatment of what selection bias is and how randomization solves that problem, see Shadish et al. 2002).

The second type of bias (reflected in publications) can result from two sources—first, when studies that only show significant results get published in academic journals and others remain in file cabinets (this is also referred as the "file drawer problem"), and second, when a study is conducted that supports prior beliefs and therefore presents statistically significant results. The first type of publication bias can also occur in randomized control trials. But the second type of bias is more likely in retrospective, non-randomized studies. This second type of publication bias is very similar to the approach commonly referred as "cherry-picking," whereby only successful projects/programs are selected for impact assessment (thus, there is a publication bias toward successful interventions). This basically implies that evaluations themselves are self-selected. In theory, the way to minimize this type of bias in retrospective studies would be to subject all projects/programs of certain size and scale to impact assessment. Alternatively, if resources to conduct evaluations are limited, a better approach would be to randomize such evaluations at an organization level. In other words, development initiatives to be evaluated in any given time period are randomly picked from a pool of all eligible candidates that meet the minimum threshold in terms of size and scale. In practice, however, it may be difficult to define "eligible" interventions given the wide differences in their types, their timings, their costs and potential for impact.

As a result, there have been a number of social experiments in developing countries. A well known example is Mexico's PROGRESA program, which provided cash transfers targeted to poor families conditional on their children attending school and obtaining health care and nutrition supplementation. Impact evaluation of this program has indicated significant gains to health, schooling, and food consumption (see Skoufias 2005 for a comprehensive overview of the design, implementation and results of the PROGRESA evaluation). Other examples related to food, agriculture, and rural development include the evaluation of the productivity impacts of a cashless microcredit program (DrumNet) in Kenya (Ashraf et al. 2008), social experiment to assess whether provision of insurance against a major source of production risk induces demand for credit to adopt new crop technologies (Giné and Yang 2008), ongoing experiments to evaluate the effectiveness of Integrated Agricultural Research for Development (IAR4D) as a new approach for conducting agricultural research and extension in Africa (FARA 2009), and the impact evaluation of the promotion of fortified orange flesh sweet potato (OFSP) on the health and nutrition indicators among children and women in rural communities in Mozambique (Arimond et al. 2007).[14]

The E and QE methods represent the emerging trends in IE and have the potential for providing rigorous proofs of development effectiveness of a publicly-supported program/policy on indicators related to poverty and food security. However, lack of external validity (incorrect inferences about whether

[14] The website of the Millennium Challenge Corporation also lists several ongoing IEs in the area of agricultural development using randomized control trials (http://www.mcc.gov/mcc/panda/activities/impactevaluation/ie-sectors/sector-agriculture.shtml).

evaluation findings would apply to different persons, times, or settings) still remains an issue with experimental designs. In addition, ethical concerns, high costs, sole focus on two average impact parameters (Intent to Treat-ITT and Average Treatment Effect on the Treated–ATT), focus on short-term effects, and impracticality of maintaining the treatment and control groups in a real world setting are often cited by critiques as limitations of social experiments. Another shortcoming, especially of E methods, is that sometimes the impacts are assessed too quickly (immediately after the project or program ends) and the long-term perspective of assessing impacts to targeted beneficiaries and beyond (including positive and negative spillover effects) is lost. This leads to the risk of prematurely applying lessons learned from the intervention to scale up efforts. These drawbacks make experimental designs unsuitable and less than a gold standard in the practice of evaluations of all types of development programs, projects, and policies (Ravallion 2009, Deaton 2007, Rodrik 2008).[15]

The limitations noted above perhaps explain why despite the advocacy for rigorous quantitative IEs based on E designs, its use in evaluation of development interventions remains uncommon. In the survey conducted by Kapoor (2002) of 78 evaluations completed by the Independent Evaluation Group of the World Bank (formerly known as Operations Evaluation Department), only one study used randomization. A more recent and broader review of the literature and meta-analysis of IEs of agricultural development interventions found that only six out of a pool of 83 selected IEs conducted since 2000 used randomized designs, and almost all of them were conducted in the last four years (IEG 2009). This review also found that most of the IEs of agricultural development interventions rely on quasi-experimental or non-experimental designs based on various innovative methods, combined with statistical matching techniques to construct a plausible counterfactual group for comparison. The use of instrumental variables (IV) to provide robustness checks or to deal with potential biases is also common.

Doing Evaluations under Real World Conditions (with Budget, Time, Data, and Political Constraints)

The growing debate between the randomistas and their critics in the recent years has also seen the rise of Real World Evaluation (RWE) designs (Figure 2). The RWE is guided by the following principles to strengthen practical IE designs: a) Basing the evaluation design on a program theory model; b) Complementing the quantitative (summative) evaluation with process evaluation; c) Incorporating contextual analysis (because many external factors along the impact pathway can influence the outcome and impact of a development effort); d) Reconstructing baseline conditions; e) The use of mixed-method approaches to strengthen validity of indicators and to improve interpretation of findings; f) Adjusting for differences between the project and comparison groups (Bamberger et al. 2006). Like the quantitative methods of IEs (E and QE), the RWE is concerned with micro-level impact assessments focused on establishing the causal link between an intervention and its effect and thus in providing the proof of project efficacy.

In summary, the development of new methods and approaches in impact assessment seen over the last 50 years, and the recent debates on the pros and cons of new and emerging evaluation approaches, point to the perennial interest among evaluators, development researchers, and policymakers in "improving the proof" and obtaining better knowledge about development effectiveness. Such debate is especially persistent for micro-level impact evaluations concerned with establishing the causal attribution. An international conference recently organized by 3ie, NONIE, AfrEA and UNICEF in Cairo was one of the attempts to bring together different voices and perspectives on impact evaluation so as to work toward

[15] Supporters of experimental designs have offered counter-arguments against some of these criticisms. For examples, Duflo et al. (2007) contend that many of the randomized evaluations conducted in recent years in developing countries have had fairly small budgets, thus making them affordable for development research. . . . Also, for large scale RCTs that cost millions of dollars (see Coalition for Evidence-Based Policy working paper, 2006, for some estimates of costs of past RCTs), attempts to measure impact using alternative methods also may have similar costs. For example, well-designed comparison group studies (QE) have data requirements that are quite similar to that of RCTs and do not necessarily offer cost savings. On the issue of the inability to assess longer-term impacts, supporters have argued that some RCTs produce valuable results in a very short time frame–within a year or two–which are a harbinger of the longer-term outcomes that are of the greatest policy significance.

a consensus (see White 2009).[16] In the following section, we attempt to bring out the implications of the evolution of and debate in IA methods for "improving the proof" of development investments focused on enhancing food security.

[16] Macro-level impact assessments and micro-level ex post impact assessments have not been included in such debates, perhaps because they are not seen by evaluators as analyses of causal attribution.

4. ASSESSMENTS OF FOOD SECURITY IMPACTS: PAST EVIDENCE, POTENTIAL FOR FUTURE AND BEST PRACTICES

Past Evidence

Although food security assessments (and famine analysis) have long been a topic of research and enquiry, the goals of achieving food security and ending poverty and hunger did not receive explicit attention in impact assessments of development interventions in the early decades. In other words, studies that address questions of how agricultural development efforts are enhancing food security of targeted populations by tracking indicators that measure "food security" were uncommon in the 1970s and 1980s. Even today, it is rare for impact assessments to include an explicit indicator called "food security" to measure the impact on this goal. Most studies measure food security impacts through changes in outcome indicators related to consumption or imply this impact through changes in outcomes related to production, income, and prices. Table 1 identifies some of the common indicators under these outcomes that are found in the impact assessment literature and lists examples of studies (focusing on the past 10 years) related to those indicators.[17] These include micro-level (Low et al. 2007, Duflo et al. 2008) as well as macro-level impact assessments (Fan et al. 2007, Evenson and Rosegrant 2003), and impact evaluations of project-specific interventions (Gilligan et al. 2008) as well as impact analysis conducted *ex post* of the scaling-up of program outputs (e.g., Evenson and Gollin 2003).

The impacts on food security and related goals (such as poverty) have been assessed at both the household or farm level, and the aggregate level. What indicators are specifically used by an impact study is a function of the objectives and program goals of an intervention (Table 1). Thus, for interventions specifically designed to address the issue of food security (food aid, social safety nets, school feeding programs, cash transfers, etc.), the indicators commonly used relate to consumption. These include food expenditures and food gap (Gilligan et al. 2008), per capita food, calorie and nutritional intake (Dillon 2008; Low et al. 2007), etc. For development activities designed to address the issue of productivity, the indicators commonly used relate to yield, production, and profitability.

Impact assessment questions are a good indicator of the objectives and scope of the development activity (or intervention) being evaluated. As exemplified in Table 1, these can be diverse depending on the size and scale of an intervention, and the timing and motivation for conducting an impact assessment. The method/approaches used for impact assessment are both a function of the impact assessment question and the type of data available. The growing practice in micro-level analysis (that focuses on specific interventions) is to use experimental or quasi-experimental designs to identify a counterfactual and attribute the change in impact indicators to the intervention (e.g., Hoddinott and Skoufias 2004, Gilligan and Hoddinnott 2007, Low et al. 2007, Dillon 2008). Such impact evaluations require specific data collection through surveys across intervention and non-intervention sites and perhaps also across multiple time periods. These types of studies have been implemented to address evaluation questions related to food security impacts at the household level.

[17] The examples and observations noted in Table 1 are samples for illustrative purposes based on author's familiarity with the literature. It is not based on a comprehensive review of the literature on impact assessments that focus on food security goals. The purpose is to highlight the diversity in methods/approaches and impact assessment questions found in the development literature related to agriculture.

Table 1: Common indicators related to food security found in the impact assessment literature and illustrative examples of methods and types of IAs

Outcomes related to food security goal	Common indicators found in the impact assessment literature	Notable observations from the literature	Illustrative examples of IA studies (since 2000) that measure some of the given outcomes and indicators		
			Study citation	Impact question related to food security outcomes	Methods and approaches used
Production	**Farm level:** Yield Unit cost of production Profitability	• One of the most common outcomes found in the assessments of many types of agricultural interventions • The impact channels between these indicators and food security are implied and indirect and not always investigated	1. Evenson and Gollin (2003) 2. Duflo et al. (2008)	1. What are the productivity impacts of improved crop germplasm research? 2. Do fertilizer and hybrid seed increase yield and profitability on small farms?	1. Mixed methods–mostly NE (impacts derived from long-term agricultural experiments and statistical analysis of secondary data and adoption surveys) 2. Experimental design applied to a community level intervention
	Aggregate level: Total Factor Productivity (TFP)		1. Thirtle et al. (2003)	1. What is the impact of research-led agric. productivity growth on poverty?	1. Econometric analysis and multi-equation models using time series data
Consumption	**HH level:** Food expenditure Food gap Total or per capita consumption Calorie intake Nutritional status measured by height and weight	• These are the most direct indicators related to food security and nutritional security • These are used to assess impacts of different types of interventions (including non-agricultural poverty reduction programs) targeted to increase household or individual's income and consumption • Micro-level IAs based on these indicators are more common	1. Hodinnott and Skoufias (2004) 2. Gilligan and Hodinnott (2007) 3. Dillon (2008) 4. Low et al. (2007)	1. What are the impacts of PROGRESA--an anti-poverty program-- on the quantity and quality of food intake? 2. What is the impact of the social safety net program on food security, consumption levels, and use of income? 3. What is the effect of increased agricultural production induced by irrigation on household consumption? 4. What are the nutritional impacts of orange fleshed sweet potatoes on young children?	1. E design based on pipeline analysis 2. QE using propensity score matching techniques 3. QE using DD and propensity score matching 4. QE using fixed effects regression model
	Aggregate level: Consumption growth Consumer surplus Food expenditures		1. Dercon et al. (2008)	1. What are the impacts of public investments in extension and roads on consumption growth and poverty in rural areas?	1. NE design based on Instrumental Variables model using Generalized Methods of Moments

Table 1. (Continued)

Outcomes related to food security goal	Common indicators found in the impact assessment literature	Notable observations from the literature	Illustrative examples of IA studies (since 2000) that measure some of the given outcomes and indicators		
			Study citation	Impact question related to food security outcomes	Methods and approaches used
Income	**HH level:** Farm income Non-farm income Total income Asset holding Gross margins **Aggregate level:** Producer surplus Number of people living below $x per day	• These are most directly linked with the poverty goal • The impact channels between these indicators and food security are implied rather than explicitly stated	1. Cocchi and Bravo-Ureta (2007) 2. Rutherford (2009) 1. Dercon et al. (2008) 2. Fan et al. (2007)	1. What is the relationship between farm income, adoption of conservation technologies and output? 2. What is the impact of BBM technology adoption on gross margins? 1. See the description above. 2. What is the impact of productivity-enhancing rice research on rural poverty in India and China?	1. QE design based on matching techniques 2. NE design based on comparison of with/without farmer household surveys 1. NE design based on Instrumental Variables model using Generalized Methods of Moments 2. Econometric analysis and models using time series data
Food prices	**Aggregate level:** Market price of food Price stability Producer surplus Consumer surplus	• The impact channels between these indicators and food security depends on the types of the groups impacted (net food consumers vs. producers) and type of commodity (food vs. non-food) • The evidence on the relationship between these indicators and food security is therefore ambiguous and not generalizable.	1. Evenson and Rosegrant (2003) 2. van den Berg and Ruben (2006)	1. How would food prices, food production and food consumption have differed if there were no investments in crop genetic improvement research in developing countries? 2. What are the food price impacts of irrigation development?	1. NE design based on econometric and GE multi-country models to estimate the counterfactual 2. NE design based on statistical analysis of before-after intervention data

Impact assessments based on existing sources of data (secondary data or primary data collected for other purposes) are usually classified as non-experimental designs and require the use sophisticated econometric techniques to establish causal attribution of an intervention (a policy change, program implementation) to the observed changes in welfare indicators derived from pre-existing surveys. Usually, such evaluations are conducted at an aggregate program level and are not specific to a time- and/or space-bound interventions (e.g., Evenson and Gollin 2003, Dercon et al. 2008). The Living Standards Measurement Studies (LSMS) sponsored by the World Bank in many developing countries have served as a rich source of data to assess the macro-level impacts of major policy changes or development programs implemented over a large scale (see Ravallion and Walle 2004 and CIEM 2008 for Vietnam; Jacoby et al. 2002 for China). These types of rich data offer great opportunities to conduct IAs and explore causal links and associations between interventions and outcome variables.

Looking Toward the Future

The traditional approaches of retrospective impact assessments based on cost–benefit and economic surplus analysis that have been widely used in agricultural development (especially to assess impacts of improved technologies) have not adequately addressed the impacts on indicators of food security (see Alston et al. 2000 and Evenson 2001 for a list of such *ex post* IAs based on economic surplus methods). At most these types of impact assessments provide estimates of consumer and producer surpluses, which can be used to infer implications on production and consumption related outcomes, especially if the analysis is focused on a food commodity. However, these estimates are at an aggregate level covering the whole domain of impact analysis (not at the household or farm level), and it can be difficult to interpret and derive actual impact on the food security or poverty status of the affected region. This inability to decipher people-related impacts has been considered one of the limitations of the traditional cost–benefit analysis framework based on the economic surplus approach. New methods are needed to incorporate other dimensions of impacts in the traditional CBA framework or to supplement the CBA analysis. As noted earlier, the extended cost–benefit analysis, which has been mainly applied to include environmental impacts, seems promising in this respect if it could be extended to include social and other non-economic impacts. For this to happen, the analysis and data collection efforts underlying steps 2 and 3 of a typical CBA-based epIA noted on page 8 need to expand to include indicators related to poverty and food security. For example, efficacy studies in step 2 (IEs) need to include impact indicators related to food security at the beneficiary levels, and data collection efforts to document the effect scale in step 3 (such as adoption studies) need to include/document profile data of beneficiaries and non-beneficiaries related to the status of food security. Such practices in steps 2 and 3, if institutionalized, can potentially provide a deeper and much broader understanding of food security related impacts that can be included as part of subsequent CBA in step 4.

Rigorous quantitative methods using experimental designs, which are the new emerging methods in the field of agricultural development, are better suited to assess multi-dimensional impacts by tracking indicators closely related to project goals. However, their potential use is restricted to impact evaluations focused on measuring development effectiveness based on the average effect of site-specific interventions. Despite its simplicity in statistical data interpretation, its utility, practicality, and application as a tool of measuring development effectiveness is questioned by many critics (for example, Deaton 2009). Moreover, one fundamental criterion of suitability for using experimental designs is that it is possible to assess causal attribution, which is possible when a number of conditions are met on how the intervention is implemented (see NONIE 2008 for a discussion on when experimental designs are best suited or not suited). Thus, E designs do not offer a magic bullet when it comes to providing the proof of development effectiveness; it is one of the many methods available in the toolkit of impact evaluators.

Based on the characteristics of an intervention and an evaluation, it is possible to define a set of scenarios with the potential for using experimental designs to establish a counterfactual (Maredia 2009). Table 2 illustrates a list (though not exhaustive) of characteristics that define an intervention and evaluation, with the spectrum of potential for using experimental designs ranging from high to low. An

intervention and evaluation that has embedded characteristics closer to the left hand side of the spectrum are in general more suitable for experimental designs than those towards the right. For the latter situations, QE, but more so NE, remain more viable options for quantitative impact evaluations. However, for scenarios where it is not possible to apply quantitative methods or they can only tell a partial impact story, qualitative methods can be used. Such possibilities exist when the interventions are characterized by activities that cut across sectors, themes or geographic areas, and when they are complex systems made up of nonlinear relationships with large numbers of unknown variables and unknown causal connections between variables.

Best Practices

Impact assessments play a vital role in generating information/knowledge about what works and what does not. However, sorting out what works or what does not in achieving outcomes related to food security and poverty goals is complex. Rarely will it be possible for a single study to trace these complex relationships in a comprehensive manner. Moreover, the strategy to answer this question may vary depending on the timeframe—what works in the short-term vs. long-term. Generation of robust knowledge that informs developmental policies and investment decisions requires a hierarchical and incremental approach to "improving the proof" through rigorous and a variety of impact assessment methods applied incrementally at the project, program, and system level. The principles of the hierarchical approach are suggested in the following four steps:

Table 2: The potential for using experimental designs in impact evaluation of a development intervention

	Potential for using experimental designs in impact evaluation		
	High		Low
Characteristics of the intervention			
Type of intervention	Single strand initiatives dealing with non-social environment		Comprehensive programs that cut across sectors, themes, and geographic areas
Objectives	Simple/explicit/short term		Complex/higher level/longer term/multi-dimensional
Geographic focus	Location specific		Multi-locations across countries and regions
Impact measured at	Individual beneficiary level	Community level	Aggregate system level
Assigning/restricting intervention to potential beneficiaries	Feasible without any ethical or political concerns		Not practical and infeasible on the grounds of ethics and political sensitivity
Characteristics of Evaluation			
Time period of planning and implementing the evaluation	Prior to the design of an intervention	After the intervention is designed but not yet implemented	After the intervention is planned and implemented
Motivation for doing an impact evaluation	Test the efficacy and effectiveness of a program as reflected in the "average effect"		Estimate the total effect of a project/program after it has been scaled-up
	Test "internal validity"		Test "external validity"
Unit of impact analysis	Project	Program	System

Source: Maredia (2009)

1. Improve the proof between project inputs, outputs and potential outcomes. As a first step toward the hierarchical approach to improving the proof, it is essential: a) that all investment decisions are based on a strong proposal, a conceptual framework, and a planning document that clearly lays out the impact pathway with clearly identified links between activities/inputs, outputs and projected outcomes; and b) that all projects have an integrated M&E plan to monitor outputs and progress toward outcomes. Whether such an M&E plan is based on classical tools of LFA, ZOPP, or the more recent participatory methods should be left to project planners and implementers. The important thing is that any method that is used should provide information to establish the link between inputs and outputs and expected outcomes.
2. Improve the proof of development outcomes realized at the beneficiary level as a result of outputs generated by development efforts. Unlike the M&E assessments proposed for all interventions in hierarchy 1, this type of assessment is suggested for a sub-set of development activities with common project outputs that collectively contribute toward a common indicator related to food security. Thus, for a selected (randomly, if feasible) and possibly representative sample of development activities that lead to similar types of project outputs, there should be periodic documentation of development outcomes at the beneficiary level using rigorous methods. Ideally, such IAs should be planned and coordinated across major development investors to ensure a common assessment framework and methodology to derive generalizable results. The literature provides guidelines for good practices in assessing impacts of outcomes at the beneficiary level (IEG 2009). This can be summarized under the categories of methods and robustness of results:

   Best practices for methodology:
   - Estimating counterfactual that minimizes selection bias.
   - Controlling for pre- and post-program differences in participants.
   - Collecting policy-relevant data at baseline and follow-up to estimate program impacts.
   - Allowing sufficient time frame for program impacts.
   - Incorporating qualitative techniques to allow for the triangulation of findings.
   - Assessing impacts and reporting results beyond the "mean outcomes."

   Best practices for robustness of results:
   - Using more than one technique to infer patterns of impact from data collected.
   - Ensuring that the treatment and comparison groups are of sufficient sizes to establish statistical inferences with minimal attrition.

Despite the claims of advocates of different methods (qualitative, E, QE, etc.) discussed in the previous section, no single method dominates across all the criteria of best practices. As suggested by Ravallion (2008), rigorous, policy-relevant evaluations should be open-minded about methodology, adapting to the problem setting and data constraints. Depending on resources available, a two-step approach can also be used toward building the evidence of development effectiveness at the beneficiary level, where as a first step, low-cost, less rigorous studies of a wide range of interventions are conducted to identify areas where an additional research investment, using more rigorous methods, is warranted. The value of this filtering step would be to generate hypotheses about what works that merits confirmation in more rigorous studies (Ravallion 2008).

1. Periodically synthesize results across hierarchical 2 studies and assessments. Such syntheses should establish the link between generic categories of project/program outputs and common shorter-term developmental outcomes (for example, increased production, income, consumption, food prices). A recent attempt to conduct statistical meta-analysis of IEs that

assess agricultural productivity impacts is one of the few examples of hierarchical 3 type assessments (IEG 2009). However, one of the limitations faced by this study in applying the methodology of statistical MA, which is commonly used in other fields, was insufficient number of observations with common, comparable measure of effect to derive statistically meaningful results. Concerted and coordinated efforts to conduct impact assessments in hierarchical 2 studies that adhere to best practices in methods and robustness of results and that use common indicators of program effectiveness can potentially lead to a large pool of studies over time to conduct statistical meta-analysis to derive generalizable patterns on what works and what does not in development.[18] As indicated before, such meta-analyses are a standard practice in other fields of public sector investments (education, health, social welfare) where the application of rigorous quantitative methods based on best practices to establish the causal link between an intervention and its outcome(s) are common in impact evaluation of social programs. Also, protocols exist on documenting evidence of such IEs from published and unpublished sources, undertaking systematic reviews, and synthesizing the results across all the documented evidence (for instance, what works clearinghouse and Campbell Collaboration). The experience gained in other fields in MA and other types of systematic syntheses can serve as a source of learning about the best practices for the agricultural development community concerned with the question of what works and what does not.

2. Document longer-term developmental impacts realized across the beneficiaries (and non-beneficiaries) as a result of scaling-up generic types of developmental outputs (such as yield enhancing technologies and long-term impacts on level and distribution of income and food consumption effects). Strategic *ex post* impact assessments of program-level outputs and macro-level assessments across programs and sectors should be conducted periodically to gain insights on the longer-term impacts on broader developmental goals. How the subjects of such assessments are selected have bearing on whether they contribute towards "providing the proof" or learning about what worked and what failed. In the case of *ex post* impact assessments, the common approach has been to "cherry-pick," whereby only successful projects/programs are selected for impact assessment. This is because in practice (given the limited resources) impact assessments focused on longer-term impacts are conducted with only the motivation to prove successes rather than to learn from failures. Thus *ex post* impact assessments of "dry holes" are rare in the agricultural development literature. This tradition will need to change if "improving the proof" by learning from both successes and failures is the main underlying goal toward a vibrant impact assessment culture in agriculture.

The vision for the hierarchical approach suggested above is that the knowledge gained from proofs of development effectiveness from hierarchy 2, 3, and 4 studies will contribute toward the development of "theory of change" that can feed in to the development of future interventions and the M&E plan in hierarchy 1.

18 To ensure that hierarchical 3 analyses lead to such knowledge and information requires that evaluations in hierarchy 2 are not self-selected (i.e. only those interventions that have high chances of success are subjected to IE) and that all the evaluation results (positive, neutral or negative) are in the public domain (i.e., published in one form or another to minimize publication bias in meta-syntheses). In theory, the way to minimize the self-selection bias would be to subject all projects/programs of certain size and scale to hierarchical 2 types of impact assessment. Alternatively, if resources to conduct evaluation are limited, a better approach would be to randomize such IEs at a program/system/investor level (e.g., across all the World Bank or OECD funded projects). In other words, development programs to be evaluated in any given time period are randomly picked from a pool of all eligible candidates that meet the minimum threshold in terms of size and scale. There are certainly challenges; but more thought needs to be devoted on how to make this suggestion practically and politically feasible to implement.

5. CONCLUSIONS AND EMERGING LESSONS IN THE PRACTICE OF IMPACT ASSESSMENT OF AGRICULTURAL DEVELOPMENT

Researchers have always conducted research on impacts of development programs, projects, and policies. Open any development-oriented academic journal or even popular media publications such as newspapers and magazines, and one will find ample stories on impacts of policy A in country X or program B in community Y. Assessing, understanding, analyzing, and evaluating impacts of public programs, projects, policies, and activities has captured the interest and attention of all the stakeholders involved. "Impact assessment" as a distinct field of enquiry on which many of these impact stories are based is defined in this paper as an *assessment of changes in indicators of program goal(s) that can be attributed to a particular intervention.* It is the analysis that establishes the causal link between an intervention and outcome(s), and sometimes to analyses that go beyond establishing the causal link by estimating the total effects vis-à-vis total costs. Depending on their timing and the level of aggregation when they occur, impact assessments can serve as a vital tool in determining how investments in a development activity can be most efficiently and equitably made to achieve program goals such as enhancing food security, reducing poverty, or stimulating economic growth.

As shown in this paper, the theory and practice of impact assessment as applied to agricultural development (including research, extension, infrastructure, marketing, policy, and other types of interventions in the area of agriculture) has evolved over the past few decades in response to changing themes in development, methodological advancements, and demand for documenting rigorous evidence of impacts emanating from development interventions.

The practice of including impact evaluations as part of the design and implementation of development programs is now the new and emerging trend (see for example the impact evaluation strategies of the Millennium Challenge Corporation and the Bill and Melinda Gates Foundation, and the new initiative by the World Bank called Agricultural Adaptations, or "AADAPT"). This has given rise to the application of experimental and quasi-experimental methods in assessing impacts on a wide variety of project specific indicators, including food security, poverty, and other non-economic and non-traditional measures of project outcomes and impacts. The growing emphasis on rigorous quantitative methods in recent years has gone hand-in-hand with increasing emphasis on mixed method, real world evaluation approaches based on less rigorous study designs. Such methods also serve useful purposes, for example in generating hypotheses about what works that merits confirmation in more rigorous studies.

The traditional approach toward accounting for impacts of a new technology that increases agricultural productivity and availability of food has been to use a cost–benefit analysis framework, where benefits are estimated based on primary and secondary data spanning many time periods and geographic areas, assumptions about the underlying relationships between model parameters, and then an estimation of economic rates of return (ROR) on project- or program-specific investments. Also, impact assessment based on success stories has been a common practice, especially in the agricultural research impact assessment literature. These approaches and methods played an important function in meeting the demand for accountability and "strategic validation" from donor communities. However, with the increasing emphasis on learning from impact assessment so as to improve the design and implementation of agricultural development programs, the value of "impact assessment" as occurring several years after the end of an intervention and only focusing on success stories is questionable. A hierarchical and incremental approach to impact assessment that integrates M&E, IE, MA, epIA and macro-level IA in hierarchical steps is suggested as a best practice guideline for making evidence-based policies and investment decisions.

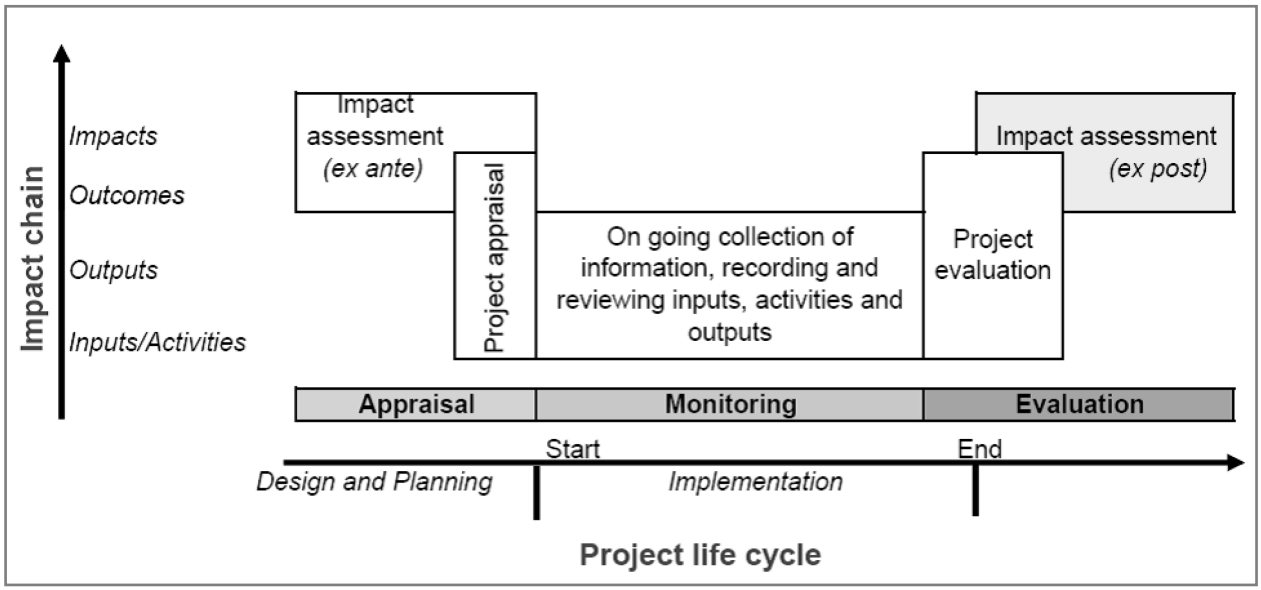
Subjecting as many development interventions as resources allow to rigorous impact assessment based on a common impact assessment framework (a common structure and model relationships, indicators of measured impacts, method of analysis, etc.) can help build a body of evidence on impacts of development interventions. After a critical mass of evidence is accumulated, these can be subjected to meta-analyses to help assimilate results across different impact assessments and build a knowledge base

on what works and what does not and why in different contexts. This can then serve as the foundation toward building more efficient and effective programs. Such a common framework is urgently needed in the field of agricultural development, and could be organized around broad development goals such as poverty, food security, and environmental sustainability. What such a common framework will look like and the organizational and coordination challenges to implementing this across development interventions funded by different donor agencies and ministries needs to be further explored.

ANNEX 1. IMPACT ASSESSMENT IN THE CONTEXT OF PROJECT CYCLE

The impact chain provides a useful way of conceptualizing the cause-and-effect relationship between different types of changes, with impact assessment focusing mainly on the changes occurring at the outcome and impact level. Several concepts and terminologies found in the literature are closely associated with impact assessment or belong to the family of monitoring, evaluation, and impact assessment. These are introduced in Figure A1, which combines the concept of impact chain with the timeline of project cycle to provide the classical view of the differences among appraisal, monitoring, and evaluation, and illustrates where impact assessment fits in this two-dimensional view of the world. In the overview of this paper, the following distinction is made between monitoring, project evaluation,[19] and impact assessment (Roche 1999).

Figure A1: A generalized view of the types of assessments in a project life cycle and impact chain framework



Timing: Monitoring occurs frequently throughout the implementation phase, and project evaluation occurs periodically (such as mid-term or end of the project). *Ex post* impact assessment as a type of evaluation occurs infrequently, usually toward or after the end of an intervention. *Ex ante* impact assessment as a tool for planning and appraisal occurs before the start of an intervention.

Analytical level: Monitoring is mainly descriptive, project evaluation is more analytical and examines processes and short-term/intermediate outcomes, and impact assessment is mainly analytical and concerned with longer-term outcomes and impacts.

Specificity: Monitoring is very specific and compares a particular plan and its results. Project evaluation does the same but also looks at processes, whereas impact assessment is less specific and in addition considers external influences and events.

[19] The term project evaluation as used in this paper is distinct from impact evaluation. It is also referred by some as "performance evaluations" (e.g., Frey and Osterloh 2006).

Conducted by: Monitoring is normally regarded as an internal function performed by project management. In contrast, project evaluation and impact assessment are generally performed (or commissioned) by funders and/or relevant ministries (with or without project management).[20]

Data source: Monitoring is based on periodic data collected by the project staff. Project evaluation draws on the data used in the monitoring system, but goes beyond to determine effects. In principle, impact assessment (*ex post*) would entail the collection of data external to the project, either in space and/or time, so that the counterfactual — the question of what would have happened without the project — may be posed, and the changes attributable to the project distinguished from those which would have arisen independently.

Purpose: By nature, monitoring activities are best suited for institutional learning for project management. Impact assessment (*ex post*) better fulfills the accountability function of evaluation and assessment (accounting for project outcomes and impacts to justify inputs). Given the timing of when they occur, impact assessments (*ex post*) have limited potential for direct feedback and learning for project management. They are, rather, designed to yield strategic lessons which may assist in the planning and execution of further similar initiatives, which will be undertaken at other locations. On the other hand, the nature of learning and accountability that occurs under project evaluation is distinct from monitoring and retrospective impact assessment. Project evaluations serve the purpose of promoting efficient quality control, efficient allocation of resources, performance (for on-going projects/programs), and improved transparency of activity to stakeholders.

*Frequency: C*ompared to appraisal and monitoring assessments, project evaluations and impact assessments are resource intensive. By contrast with monitoring and project evaluations, which ideally should always be performed, *ex post* impact assessments are generally conducted for only a relatively small subset of the total range of activities funded by an agency.

[20] Most international aid agencies and organizations (the World Bank, OECD, IADB, etc.) have independent evaluation units solely responsible for evaluation, which includes project/program evaluation as well as impact assessment. In this context, impact assessment is considered one type of evaluation.

ANNEX 2. DOMINANT PARADIGMS, SEQUENTIAL THEMES, AND EVOLUTION OF POPULAR IDEAS IN AGRICULTURAL DEVELOPMENT, 1950S-2000S

Table A2: Dominant paradigms, sequential themes, and evolution of popular ideas in agricultural and rural development, 1950s–2000s

	1950s	1960s	1970s	1980s	1990s	2000s
Dominant paradigms	Modernization, dual economy					
		Raising yields on efficient small farms				
				Process, participation, empowerment		
					Sustainable livelihoods approach	
Popular rural development emphases	Community development					
		Small farm growth				
			Integrated rural development			
				Market liberalization		
						Poverty reduction strategies
Rural development ideas	**1950s** Modernization Dual economy model "backward" agriculture Community development Lazy farmer	**1960s** Transformation approach Technology transfer Mechanization Agricultural extension Growth role of agriculture Green revolution (start) Rational peasants	**1970s** Redistribution with growth Basic needs Integrated rural development State agricultural policies State-led credit Urban bias Induced innovation Green revolution (cont'd) Rural growth linkages	**1980s** Structural adjustments Free markets "getting prices right" Retreat of the state Rise of NGOs Farming systems research (FSR) Food security and famine analysis Rural development as process not product Women in development (WID) Poverty alleviation	**1990s** Microcredit Participatory approaches Actor-oriented rural development Stakeholder analysis Rural safety nets Gender and development Environment and sustainability Poverty reduction	**2000s** Sustainable livelihoods Innovation systems Millennium Development Goals Evidence-based policymaking Integrated Natural Resource Management Participation for results Decentralization Sector-wide approaches Social protection Poverty eradication

Source: Adapted from Ellis and Biggs (2001)

ANNEX 3. QUANTITATIVE EVALUATION DESIGNS CONCERNED WITH ESTABLISHING COUNTERFACTUAL

Experimental designs (E) are based on a lottery system of randomly allocating the intervention among eligible beneficiaries. The random assignment process itself creates comparable treatment and control groups that are statistically equivalent to one another. This is considered a powerful approach because, in theory, a control group generated through random assignment serves as a perfect counterfactual, free from the troublesome selection bias issues that often plague evaluations (see Box 3). Only rigorous experimental studies can definitively establish that a program *causes* changes in outcomes.

Quasi-experimental studies (QE) also examine outcomes; however, they do not involve randomly assigning participants to treatment and control groups, *a priori*. A quasi-experimental study compares outcomes for individuals receiving program activities with outcomes for a similar group of individuals not receiving program activities using econometric and statistical techniques to estimate propensity score matching, matched difference in differences estimators, etc. It also includes regression discontinuity techniques to estimate program effects. These evaluation designs draw on a growing theoretical and empirical literature on the estimation of program effects from non-experimental data (e.g., Heckman et al. 1998, Smith and Todd 2005, Gilligan and Hoddinott 2007).

Non-experimental statistical designs (NE) have traditionally been the most widely used approaches for assessing program effects (Figure 2). They are used when it is not possible to randomly select a control group or identify a suitable comparison group through matching methods. In such situations, program participants are compared to non-participants using one of the following approaches, some more rigorous than others. A more rigorous approach under this category is the use of instrumental variables (IV) in a statistical analysis (e.g., Jayne et al. 2004). Good instruments help predict program participation and allow the researcher to predict the outcome of program participants and non-participants. Other methods often used in retrospective IAs include extrapolating/inferring the difference in an impact indicator between the two groups based on results of other controlled experiments (for example, predicting the difference between an improved and traditional technology based on research trials); econometric simulation models based on secondary data to construct a "without" intervention scenario (such as the use of IMPACT model by Evenson and Rosegrant, 2003), or simply comparing the participant and non-participant groups that are formed ex-post of project scale up (such as adopters vs. non-adopters of a technology—e.g., Dey et al. 2006) and acknowledging the selection bias implicit in the results.

The E and QE methods can be combined with comparison before and after the group's involvement in a program (known as "pre/post" or "reflexive" designs). Impact estimates based on such longitudinal data involve statistical procedures known as "difference-in-difference" or double difference (DD) estimators.

The comparison of design elements and pros and cons of the three basic quantitative designs is given in Table A3.

Table A3: Comparison of different quantitative evaluation designs

Design elements	Quantitative Methods		
	Experimental	Quasi-experimental	Non-experimental
Treatment group (Group that receives the intervention)	Assigned randomly from equally eligible and willing participants	Identified based on some eligibility criteria set by the project implementers, or they may be a naturally occurring group (self-selected, non-randomized)	Identified based on some eligibility criteria set by the project implementers, or they may be a naturally occurring group (self-selected, non-randomized)
Comparison / control group (Group that does not receive the intervention)	Assigned randomly from equally eligible and willing participants	Comparison group is identified using matching techniques, reflexive comparisons (before/after) or regression discontinuity technique panel data.	Program participants are compared to non-participants using statistical methods to account for differences between the two groups and to reduce selection bias (e.g., instrumental variables) or other methods (e.g., extrapolating data from other experiments or simply comparing the naturally occurring groups).
Intervention setting	Controlled	Controlled or not-controlled	Not relevant for this design
Timing	Evaluation must be planned and implemented as part of the intervention from the beginning (*ex ante*). Can be combined with pre-post designs (before/after)	Evaluation design can be independent of project/ program design. It is possible to conduct a sound evaluation using this design after the project has begun and/or if only post-intervention data are available. It can be combined with pre-post designs	Non-experimental designs are independent of project/program design and implementation. They usually occur after the project/program has begun or is fully implemented. It is based on existing data sources. Pre-post design in not feasible.
Advantages	It is free from the troublesome selection bias issues that exist in all evaluations. Simplicity in interpreting results.	It can draw on existing data sources and is thus often quicker and cheaper to implement. It is free from ethical concerns.	It is relatively easy to implement (economically and politically) because it can draw on existing data sources. It is free from ethical concerns.
Drawbacks	It may be expensive, time consuming, politically difficult, and unethical to implement; risk of contamination of control group; difficult to ensure that assignment is truly random.	The reliability of the results is often reduced, as the methodology may not completely solve the problem of selection bias. Some of the techniques can be statistically complex and may require expertise in the design of the evaluation and in analysis and interpretation of the results.	Reliability of results is reduced as the methodology is less robust statistically. Statistical complexities may require some expertise in the design of the evaluation and in the analysis and interpretation of results. Full correction of selection bias remains a challenge.

Source: Various overviews found in the literature (e.g., Baker 2000, Ravallion 2008, Maredia 2009)

REFERENCES

Adato, M., and R. Meinzen-Dick, eds. 2007. *Agricultural research, livelihoods and poverty: Studies of economic and social impact in six countries*. Washington, D.C. and Baltimore, Md.: International Food Policy Research Institute (IFPRI) and John Hopkins University Press.

Alston, J. M., G. W. Norton, and P. G. Pardey. 1998. *Science under scarcity: Principles and practice for agricultural research evaluation and priority setting*. Wallingford, U.K.: CAB International.

Alston, J. M., C. Chan-Kang, M. C. Marra, P. G. Pardey, and T. J. Wyatt. 2000. A meta-analysis of rates of return to agricultural R&D: Ex pede Herculem? IFPRI Research Report 113. Washington, D.C.: International Food Policy Research Institute (IFPRI).

Arimond, M., A. de Brauw, D. Gilligan, C. Hotz, R. Labarta, C. Loechl, and J. V. Meenakshi. 2007. Baseline survey report of OFSP evaluation. HarvestPlus Challenge Program, IFPRI, Washington, D.C. Photocopy.

Ashraf, N., X. Gine, and D. Karlan. 2008. Finding missing markets (and a disturbing epilogue): Evidence from an export crop adoption and marketing intervention in Kenya. Working Paper No. 08-065. Boston, Mass., U.S.A.: Harvard Business School. Also, published as World Bank Policy Research Working Paper 4477, Washington, D.C.

Bailar, J. C. 1997. The promise and problems of meta-analysis. *The New England Journal of Medicine* 337 (8): 559–561.

Baker, J. L. 2000. Evaluating the impact of development projects on poverty: A handbook for practitioners. Washington, D.C.: The World Bank.

Bamberger, M., and J. Rugh. 2008. Real world evaluation—Working under budget, time, data, and political constraints: Overview. Prepared for the American Evaluation Association Professional Development Workshop Session 21, November 5.

Bamberger, M., J. Rugh, and L. Mabry. 2006. RealWorld evaluation: Working under budget, time, data, and political constraints. Newbury Park, Calif., U.S.A.: Sage Publications.

Banerjee, A. V., A. H. Amsdan, R. H. Bates, J. Bhagwati, A. Deaton, and N. Stern. 2007. Making aid work. Cambridge, Mass., U.S.A.: The MIT Press.

Bennett, J. 2009. Advancing ex-post impact assessment of environmental and social impacts of CGIAR research. Paper submitted to the Standing Panel on Impact Assessment of the CGIAR Science Council.

Berman, N. G., and R. A. Parker. 2002. Meta-analysis: Neither quick nor easy. *BMC Medical Research Methodology* 2: 10. Available online <http://www.biomedcentral.com/1471-2288/2/10>.

Catley, A., J. Burns, D. Abebe, and O. Suji. 2007. Participatory impact assessment: A guide for practitioners. Medford, Mass., U.S.A.: Feinstein International Center, Tufts University. <http://www.reliefweb.int/rw/lib.nsf/db900sid/SHIG-7L2K8C/$file/Tufts_Oct2008.pdf?openelement>. Accessed in June 2009.

Center for Global Development (CGD). 2006. *When will we ever learn? Improving lives through impact evaluation*. Washington, D.C.: Center for Global Development.

Central Institute for Economic Management (CIEM). 2008. Final report of the impact evaluation of the Northern Mountains Poverty Reduction Project (NMPRP). Submitted to the World Bank, Hanoi, Vietnam, June.

Chambers, R. 2007. Poverty research: Methodologies, mindsets and multidimensionality. IDS Working Paper 293. Brighton, U.K.: Institute of Development Studies at the University of Sussex.

_____. 1994. The origins and practice of participatory rural appraisal. *World Development* 27 (7): 953–969.

_____. 1992. *Rural appraisal: Rapid, relaxed, and participatory*. IDS Discussion Paper 311. Brighton, U.K.: Institute of Development Studies at the University of Sussex.

Coalition for Evidence-Based Policy. 2006. Which study designs can produce rigorous evidence of program effectiveness? A brief overview. Working Paper, Coalition for Evidence-Based Policy, Washington, D.C.

<http://www.evidencebasedpolicy.org/docs/RCTs_first_then_match_c-g_studies-FINAL.pdf>. Accessed in August 2009.

Cocchi, H., and B. E. Bravo-Ureta. 2007. On-site costs and benefits of soil conservation among hillside farmers in El Salvador. Working Paper 4/07, Office of Evaluation and Oversight (OVE). Washington, D.C.: Inter-American Development Bank.

Deak, A., J. Pettit, P. Taylor, and I. Vogel. 2006. Learning for social change: exploring concepts, methods and practice. Brighton, U.K.: Institute of Development Studies.

Deaton, A. 2007. Evidence-based aid must not become the latest in a long string of development fads. In *Making Aid Work*, ed. A. V. Banerjee, A. H. Amsden, R. H. Bates, J. Bhagwati, A. Deaton, and N. Stern. Cambridge, Mass., U.S.A.: The MIT Press.

_____. 2009. Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. NBER Working Paper 14690. Cambridge, Mass., U.S.A.: National Bureau for Economic Research.

Dercon, S., D. O. Gilligan, J. Hoddinott, and T. Woldehanna. 2008. The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. Discussion Paper 840. Washington, D.C.: International Food Policy Research Institute (IFPRI).

Dey, M. M., P. Kambewa, M. Prein, D. Jamu, F. J. Paraguas, and R. M. Briones. 2007. Impact of the development and dissemination of integrated aquaculture-agriculture (IAA) technologies in Malawi. In *International research on natural resource management: Advances in impact assessment*, ed. H. Waibel and D. Zilberman. Wallingford, U.K.: CAB International.

Dillon, A. 2008. Access to irrigation and the escape from poverty: Evidence from northern Mali. Discussion Paper 782. Washington, D.C.: International Food Policy Research Institute (IFPRI).

Douthwaite, B., B. S. Alvarez, S. Cook, R. Davies, P. George, J. Howell, R. Mackay, and J. Rubiano. 2007. The impact pathways approach: A practical application of program theory in research-for-development. *Canadian Journal of Program Evaluation* 22 (2): 127–159.

Duflo, E., R. Glennerster, and M. Kremer. 2007. Using randomization in development economics research: A toolkit. CEPR Discussion Paper 6059. London, U.K.: Center for Economic Policy Research.

Duflo, E., M. Kremer, and J. Robinson. 2008. How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *American Economics Review* 98 (2): 482–488.

Duflo, E. 2006. Field experiments in development economics. Prepared for the World Congress of the Econometric Society, Department of Economics and Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology, January 2006.

Ellis, F., and S. Biggs. 2001. Evolving themes in rural development 1950s–2000s. *Development Policy Review* 19 (4): 437–448.

Espey, M., J. Espey, and W. D. Shaw. 1994. Price elasticity of residential demand for water: a meta-analysis. *Water Resources* 33: 1369–1374.

Evenson, R. E. 2001. Economic impacts of agricultural research and extension. In *Handbook of agricultural economics*, vol. 1A, ed. B. L. Gardner and G. C. Rausser. Amsterdam, the Netherlands: Elsevier.

Evenson, R. E., and D. Gollin, eds. 2003. *Crop variety improvement and its effect on productivity: The impact of international agricultural research.* Wallingford, U.K.: CAB International.

Evenson, R. E., and Y. Kislev. 1975. Agricultural research and productivity. New Haven, Conn., U.S.A.: Yale University Press.

Evenson, R. E., and M. W. Rosegrant. 2003. The economic consequences of CGI programs. In *Crop variety improvement and its effect on productivity*: *The impact of international agricultural research*, ed. R. E. Evenson and D. Gollin. Wallingford, U.K.: CAB International.

Fan, S., C. Chan-Kang, K. Qian, and K. Krishnaiah. 2007. National and international agricultural research and rural poverty: The case of rice research in India and China. In *Agricultural research, livelihoods, and poverty: Studies of economic and social impact in six countries*, ed. M. Adato and R. Meinzen-Dick. Washington, D.C. and Baltimore, Md.: International Food Policy Research Institute (IFPRI) and John Hopkins University Press.

Fan, S. 2007. Agricultural research and urban poverty in China and India. In *Agricultural research, livelihoods, and poverty: Studies of economic and social impacts in six countries*, ed. M. Adato and R. Meinzen-Dick. Washington, D.C. and Baltimore, Md.: International Food Policy Research Institute (IFPRI) and Johns Hopkins University Press.

FAO (Food and Agriculture Organization of the United Nations). 2003. *Trade reforms and food security: Conceptualizing the linkages*. Rome: Commodity Policy and Projections Service, Commodities and Trade Division, FAO.

_____. 2005. Summary of world food and agricultural statistics. Rome: FAO. <http://www.fao.org/es/ess/sumfas/sumfas_en_web.pdf>. Accessed in June 2009.

FARA (Forum for Agricultural Research in Africa). 2009. Sub-Saharan Africa challenge programme: Research plan and programme for impact assessment. FARA, Accra, Ghana.

Fisher, R. A. 1935. *The design of experiments*. New York: Hafner Press.

Frey, B. S., and M. Osterloh. 2006. Evaluations: Hidden costs, questionable benefits, and superior alternatives. Working Paper No. 2006-23. Basel, Switzerland: CREMA Gellertstrasse.

Gilligan, D., and J. Hoddinott. 2007. Is there persistence in the impact of emergency food aid? Evidence on consumption, food security, and assets in rural Ethiopia. *American Journal of Agricultural Economics* 89 (2): 225–242.

Gilligan, D., J. Hoddinott, and A. Taffesse. 2008. The impact of Ethiopia's productive safety net programme and its linkages. International Food Policy Research Institute (IFPRI), Washington, D.C. Photocopy.

Giné, X., and D. Yang. 2009. Insurance, credit, and technology adoption: Field experimental evidence from Malawi. *Journal of Development Economics* 89 (1): 1–11.

Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3–8.

Goldstein, M., E. Skoufias, and A. Fiszbein. 2008. Assessing our work on impact evaluation. PREMnotes on Poverty No. 126. Washington, D.C.: The World Bank.

Griliches, Z. 1958. Research costs and social returns: hybrid corn and related innovations. *Journal of Political Economy* 66: 419–431.

_____. 1963. The sources of measured productivity growth: U.S. agriculture 1940–1960. *Journal of Political Economy* 74 (4): 331–346.

_____. 1964. Research expenditures, education, and the aggregate agricultural production function. *American Economic Review* 54 (6): 961–974.

GTZ. 1988. ZOPP (an introduction to the method). Eschborn, Germany: Deutsche Gesellschaft fur Technische Zusammenarbeit.

_____. 1998. ZOPP: Objectives-oriented project planning—A planning guide for new and ongoing projects and programmes. Eschborn, Germany: Deutsche Gesellschaft fur Technische Zusammenarbeit.

Guijt, I., J. Gaventa, D. A. Johnson, J. Blauert, and M. Estrella. 2000. Learning from change: Issues and experiences in participatory monitoring and evaluation. Ottawa, Canada: International Development Research Centre/Practical Action Publishing.

Guijt, I. 2007. Assessing and learning for social change: A discussion paper. Brighton, U.K.: Institute of Development Studies of the University of Sussex.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66 (5): 1017–1098.

Hoddinott, J., and E. Skoufias. 2004. The impact of PROGRESA on food consumption. *Economic Development and Cultural Change* 53 (1): 37–61.

Görg, H., and E. Strobl. 2001. Multinational companies and productivity spillovers: A meta-analysis. *Royal Economic Society* 111 (475): F723–F739.

Howes, M. 1992. Linking paradigms and practice: Key issues in the appraisal, monitoring and evaluation of British NGO projects. *Journal of International Development* 4 (4): 375–396.

IFAD. 2003. IFAD evaluation policy. Prepared for the International Fund for Agricultural Development, Executive Board's Seventy-Eighth Session, Rome, April 9–10, 2003.

Independent Evaluation Group (IEG). 2009. Measuring the impacts of agricultural projects: A meta-analysis and three impact evaluations. IEG of the World Bank, Washington, D.C. Photocopy.

Jacoby, H., G. Li, and S. Rozelle. 2002. Hazards of expropriation: Tenure insecurity and investment in rural China. *American Economic Review* 92 (5): 1420–1447.

Jayne, T. S., T. Yamano, and J. Nyoro. 2004. Interlinked credit and farm intensification: Evidence from Kenya. *Agricultural Economics* 31: 209–218.

Jensen, R. 2007. The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *The Quarterly Journal of Economics* 122 (3): 879–924.

Kapoor, A. G. 2002. *Review of impact evaluation methodologies used by the Operations Evaluation Department over past 25 years.* Washington, D.C.: OED (now IEG), World Bank.

Kelley, T., J. Ryan, and H. Gregersen. 2008. Enhancing *ex post* impact assessment of agricultural research: The CGIAR experience. *Research Evaluation* 17 (3): 201–212.

Kerr, J., and K. Chung. 2001. Evaluation watershed management projects. CAPRi Working Paper No. 17. Washington, D.C.: International Food Policy Research Institute (IFPRI).

La Rovere, R., D. Flores, P. Aquino, and J. Dixon. 2008. Through the livelihoods lens: Lessons from Mexico and Nepal on integrating approaches and metrics for impact assessment. Paper presented at the workshop on "Rethinking Impact: Understanding the Complexity of Poverty and Change," Cali, Colombia, March 26–28, 2008.

Low, J., M. Arimond, N. Osman, B. Cunguara, F. Zano, and D. Tschirley. 2007. A food-based approach introducing orange fleshed sweet potatoes, increased vitamin A intake and serum retinol concentrations in young children in rural Mozambique. *Journal of Nutrition* 137 (5): 1320–1327.

Mancini, F., A. H. Van Bruggen, and J. L. S. Jiggins. 2007. Evaluating cotton integrated pest management farmer field school outcomes using the sustainable livelihoods approach in India. *Experimental Agriculture* 43: 97–112.

Maredia, M. K., and D. A. Raitzer. 2006. *CGIAR and NARS partner research in Sub-Saharan Africa: Evidence of impact to date*. Rome: CGIAR Science Council Secretariat.

Maredia, M. K. 2009. The scope and feasibility of using experimental designs in evaluating impacts of investments in agricultural research for development. Prepared for the Standing Panel on Impact Assessment of the CGIAR Science Council, July 2009. Photocopy.

McCracken, J. A., J. N. Pretty, and G. R. Conway. 1988. *An introduction to rapid rural appraisal for agricultural development*. London: International Institute for Environment and Development.

Mohr, L. B. 1995. *Impact analysis for program evaluation*, 2nd ed. Thousand Oaks, Calif., U.S.A.: Sage Publications.

_____. 1999. The qualitative method of impact analysis. *American Journal of Evaluation* 20: 69–84.

NONIE (Network of Networks on Impact Evaluation). 2008. NONIE impact evaluation guidance. Report of the Subgroup 2, NONIE, World Bank. <http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf>. Accessed in March 2009.

Phillips, J. M. 1994. Farmer education and farmer efficiency: A meta-analysis. *Economic Development and Cultural Change* 43: 149–165.

Pingali, P. L. 2001. *Milestones in impact assessment research in the CGIAR, 1970–1999. With an annotated bibliography of impact assessment studies conducted in the CGIAR, 1970–1999, prepared by Matthew P. Feldmann.* Mexico: Standing Panel on Impact Assessment, Technical Advisory Committee of the Consultative Group on International Agricultural Research.

Ravallion, M., and D. van de Walle. 2004. Breaking up the collective farms: Welfare outcomes of Vietnam's massive land privatization. *Economics of Transition* 12 (2): 201–236.

Raitzer, D., and T. G. Kelley. 2008. Benefit-cost meta-analysis of investment in the international agricultural research centers of the CGIAR. *Agricultural Systems* 96: 108–123.

Ravallion, M. 2008. Evaluating anti-poverty programs. In *Handbook of development economics*, vol. 4, ed. H. Chenery and T. N. Srinivan. Amsterdam, The Netherlands: Elsevier.

Ravallion, M. 2009. Should the Randomistas Rule? *The Economists' Voice* 6 (2): Article 6.

Roche, C. 1999. Impact assessment for development agencies: Learning to value change. Oxford, U.K.: Oxfam.

Rodrik, D. 2008. The new development economics: We shall experiment, but how shall we learn? In *What works in development? Thinking big vs. thinking small*, ed. J. Cohen and W. Easterly. Washington, D.C.: Brookings Institution Press.

Rogers, P. J., K. Stevens, and J. Bovmal. 2009. Qualitative cost-benefit evaluation of complex, emergent programs. *Evaluation and Program Planning* 32 (1): 83–90.

Rutherford, A. S. 2008. Broad bed maker technology package innovations in Ethiopian farming systems: An *ex post* impact assessment. Research Report 20. Nairobi, Kenya: ILRI (International Livestock Research Institute).

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. Experimental and quasi-experimental designs for generalized causal inference. Boston, Mass., U.S.A.: Houghton Mifflin.

Skoufias, E. 2005. *PROGRESA and its impact on the welfare of rural households in Mexico.* Research Report 139. Washington, D.C.: International Food Policy Research Institute.

Smith, J., and P. Todd. 2005. Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics* 125(1–2): 303–353.

Taylor, P., and J. Fransman. 2004. Learning and teaching participation: Exploring the role of higher learning institutions as agents of development and social change. IDS Working Paper 209. Brighton, U.K.: Institute of Development Studies of the University of Sussex.

Theis, J., and H. Grady. 1991. *Participatory rapid appraisal for community development.* London: Save the Children Fund.

Therivel, R., and M. R. Partidario. 1996. The practice of strategic environmental assessment. London: Earthscan.

Therivel, R., E. Wilson, S. Thompson, D. Heaney, and D. Pritchard. 1992. Strategic environmental assessment. London: RSPB/Earthscan.

Thiam, A., B. E. Bravo-Ureta, and T. E. Rivas. 2001. Technical efficiency in developing country agriculture: A meta-analysis. *Agricultural Economics* 25 (2–3): 235–243.

Thirtle, C., L. Lin, and J. Piesse. 2003. The impact of research-led agricultural productivity growth on poverty reduction in Africa, Asia, and Latin America. *World Development* 31 (12): 1959–1975.

Thompson, S. G., and S. J. Pocock. 1991. Can meta-analyses be trusted? *Lancet* 338: 1127–1130.

Valadez, J., and M. Bamberger, eds. 1994. Monitoring and evaluating social programs in developing countries. Economic Development Institute of the World Bank Series. Washington, D.C.: World Bank.

van den Berg, M., and R. Ruben. 2006. Small-scale irrigation and income distribution in Ethiopia. *Journal of Development Studies* 42 (5): 868–880.

Walker, T., M. Maredia, T. Kelley, R. La Rovere, D. Templeton, G. Thiele, and B. Douthwaite. 2008. *Strategic guidance for ex post impact assessment of agricultural research.* Rome: CGIAR Science Council Secretariat.

Watts, J., R. MacKay, D. Horton, A. Hall, B. Douthwaite, R. Chambers, and A. Acosta. 2003. Institutional learning and change: An introduction. ISNAR Discussion Paper No. 3. The Hague, The Netherlands: International Service for National Agricultural Research (ISNAR).

White, H., ed. 2009. *Designing impact evaluations: Different perspectives.* International Initiative for Impact Evaluation (3ie) Working Paper No. 4. New Delhi: 3ie.

World Bank. 2007. *World development report: Agriculture for development*. Washington, D.C.: The World Bank.

Zeddies, J., R. P. Schaab, P. Neuenschwander, and H. R. Herren. 2001. Economics of biological control of cassava mealybug in Africa. *Agricultural Economics* 24: 209–219.

IFPRI DISCUSSION PAPERS
Prepared for the "Millions Fed: Proven Successes in Agricultural Development"

910. *Combating stem and leaf rust of wheat: Historical perspective, impacts, and lessons learned.* H. J. Dubin and John P. Brennan, 2009.

911. *The Asian Green Revolution*. Peter B. R. Hazell, 2009.

912. *Controlling cassava mosaic virus and cassava mealybug in Sub-Saharan Africa.* Felix Nweke, 2009.

913. *Community forestry in Nepal: A policy innovation for local livelihoods.* Hemant Ojha, Lauren Persha, and Ashwini Chhatre, 2009.

914. *Agro-environmental transformation in the Sahel: Another kind of "Green Revolution."* Chris Reij, Gray Tappan, and Melinda Smale, 2009.

915. *The case of zero-tillage technology in Argentina.* Eduardo Trigo, Eugenio Cap, Valeria Malach, and Federico Villarreal, 2009.

916. *Zero tillage in the rice-wheat systems of the Indo-Gangetic plains: A review of impacts and sustainability implications.* Olaf Erenstein, 2009.

917. *The impact of shallow tubewells and boro rice on food security in Bangladesh.* Mahabub Hossain, 2009.

918. *Hybrid rice technology development: Ensuring China's food security.* Jiming Li, Yeyun Xin, and Longping Yuan, 2009.

919. *Pearl millet and sorghum improvement in India.* Carl E. Pray and Latha Nagarajan, 2009.

920. *Institutional reform in the Burkinabè cotton sector and its impacts on incomes and food security: 1996–2006.* Jonathan Kaminski, Derek Headey, and Tanguy Bernard, 2009.

921. *Private sector responses to public investments and policy reforms: The case of fertilizer and maize market development in Kenya*. Joshua Ariga and T. S. Jayne, 2009.

922. *The mungbean transformation: Diversifying crops, defeating malnutrition.* Subramanyan Shanmugasundaram, J. D. H. Keatinge, and Jacqueline d'Arros Hughes, 2009.

923. *The global effort to eradicate rinderpest*. Peter Roeder and Karl Rich, 2009.

924. *Rural and urban linkages: Operation Flood's role in India's dairy development.* Kenda Cunningham, 2009.

925. *Rich food for poor people: Genetically improved tilapia in the Philippines*. Sivan Yosef, 2009.

926. *"Crossing the river while feeling the rocks:" Incremental land reform and its impact on rural welfare in China.* John W. Bruce and Zongmin Li, 2009.

927. *Land-tenure policy reforms: Decollectivization and the Doi Moi System in Vietnam.* Michael Kirk and Tuan Nguyen, 2009.

928. *Improving diet quality and micronutrient nutrition: Homestead food production in Bangladesh.* Lora Iannotti, Kenda Cunningham, and Marie Ruel, 2009.

929. *Improving the proof: Evolution of and emerging trends in impact assessment methods and approaches in agricultural development.* Mywish K. Maredia, 2009.

For all discussion papers, please go to www.ifpri.org/pubs/pubs.htm#dp.
All discussion papers can be downloaded free of charge.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org

IFPRI ADDIS ABABA

P. O. Box 5689
Addis Ababa, Ethiopia
Tel.: +251 11 6463215
Fax: +251 11 6462927
Email: ifpri-addisababa@cgiar.org

IFPRI NEW DELHI

CG Block, NASC Complex, PUSA
New Delhi 110-012 India
Tel.: 91 11 2584-6565
Fax: 91 11 2584-8008 / 2584-6572
Email: ifpri-newdelhi@cgiar.org