EVALUATION OF AGRICULTURAL EXTENSION AND ADVISORY SERVICES

A MEAS Training Module

By Murari Suvedi

September 2011

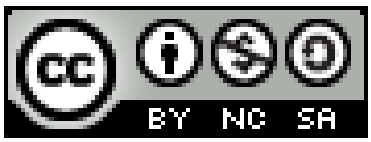
MICHIGAN STATE UNIVERSITY

USAID FROM THE AMERICAN PEOPLE

EVALUATION OF AGRICULTURAL EXTENSION AND ADVISORY SERVICES

A MEAS Training Manual

Professor Murari Suvedi
Michigan State University



This training module was produced as part of the United States Agency for International Development, USAID, project "Modernizing Extension and Advisory Services", MEAS.

www.meas-extension.org
Leader with Associates Cooperative Agreement No. AID-OAA-L-10-00003

The module was made possible by the generous support of the American people through USAID. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGEMENTS

Evaluation practitioners and international agricultural development experts have contributed to the development of this training manual by offering guidance and/or reviewing the manuscript.

Mary Andrews, Michigan State University

Kirk Heinze, Michigan State University

Diane Ruonavaara, Michigan State University

Gail Vander Stoep, Michigan State University

R. Kirby Barrick, University of Florida

Shawn Morford, Program Manager, Rural Development Initiatives, Inc.

Angela Hobson, Graduate Assistant in the Department of Community, Agriculture, Recreation and Resources Studies at Michigan State University as well as Andrea Bohn, MEAS Project Manager at the University of Illinois at Urbana-Champaign helped in editing and formatting the manual.

Resources from diverse research and evaluation literature have been utilized to develop this training manual. Samples of evaluation instruments, reports and presentation materials are drawn from author's work at Michigan State University. Specifically, this manual draws heavily from the following previous digital and print publications:

Suvedi, M., & Morford, S. (2003). Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia. Forrex-Forest Research Extension Partnership, Kamloops, B.C. Forrex Series 6.

Suvedi, M., Heinze, K., & Ruonavaara, D. (2008). How to conduct evaluation of extension programs. In Suvedi, M., Singh, B., Vijayaragavan, K., Padaria, R., & Wason, M. (Eds.) *Evaluation capacity building in rural resource management: A manual*. India: Golden Printers & Publishers.

TABLE OF CONTENTS

ABOUT THE EVALUATION TRAINING MANUAL

Planners and policy-makers in extension systems are busy professionals who may lack the skills and core competencies to guide a systematic evaluation of agricultural extension initiatives and programs. With exposure to evaluation methods and procedures, they could contribute significantly in guiding program evaluation and use results to improve future programs.

This training module is designed for policy makers, upper level administrators, and program managers in public sector institutions who have an extension mandate (including university faculty having responsibilities in extension, rural development, and food and nutrition security programs) and administrators and program officers who fund and manage projects having extension components. The purpose of this training module is to expose national level policy makers, project managers, and funding agency personnel to various models and theories of program evaluation.

Specifically, participants in this evaluation training module will be able to:

- understand program evaluation – what it is and reasons for doing it;
- describe evaluation principles and frequently used models of program evaluation;
- identify indicators of program success for a given agricultural extension project/policy;
- select appropriate methods/techniques of data collection for conducting process and impact evaluations;
- understand the use of statistics to analyze data, interpret results, write evaluation reports, and share findings with stakeholders; and
- develop evaluation plans to document impacts of extension programs.

Workshops corresponding to this manual can be offered either as a four-day training course in-country or as a regional course, based on demand. The structure and content are:

Day One: Introduce participants to the context of program evaluation in agricultural extension, evaluation concepts and principles, and frequently used models for evaluation of programs and policies (Chapters 1 and 2).

Day Two: Introduce various methods of evaluation data collection – qualitative and quantitative – including strengths and weaknesses of each method (Chapters 3-6).

Day Three: Introduce the concept of sampling, review commonly used data analysis procedures and interpretation of results, and present strategies for sharing evaluation findings with diverse stakeholders (Chapters 7-10).

Day Four: Participants develop an evaluation plan for a participant-identified extension program needing evaluation.

The content of the training manual is brief and straightforward. Some of the text is presented as bullet points for ease of reading. Power Point presentations are available on each major topic and a list of key resources is provided for further information. For illustrative purposes, samples of evaluation plans, data collection instruments, and evaluation reports are included in the Appendix. Every attempt is made to make this guide user-friendly.

CHAPTER I

History and Development of Agricultural Extension and Advisory Services

Education or learning about new methods of farming has been a function of every social system. Traditionally, farmers and fishermen learned about new methods of farming outside of the formal school system. The beginning of formal agricultural educational systems started in the United Kingdom during the middle of the nineteenth century. During the mid-19th century potato famine in Ireland, agricultural advisors taught farmers how to grow different food crops (Swanson & Claar, 1984).

A few years later, in 1867, Oxford and Cambridge universities in England started sharing the practical knowledge generated by their faculty with neighboring communities. The education programs were non-formal in nature, i.e., people did not need to be admitted into the university to take classes, no degrees or diplomas were awarded for attendance, participation was voluntary, and knowledge useful to solve a problem or issue facing the community was shared by university faculty. The term "extension" was first used to describe these applied education programs organized by these universities (Swanson & Claar, 1984).

Although several industrializing countries had some form of extension service in the later part of the 19th century, it was not formalized within public institutions. Rivera (1991) reports that Japan, in 1893, was the first country to establish by formal policy mandate a national agricultural extension system.

Higher education in agriculture in the United States started in 1855 with the establishment of Michigan Agriculture College. In the same year, Pennsylvania chartered a school at the request of the Pennsylvania State Agricultural Society. In the following years, many states established colleges of agriculture. The United States Congress passed the Morrill Act in 1862 establishing a college in every U.S. state for the common people to teach agriculture, home economics, mechanical arts and military science. Some 25 years later, in 1887, the U.S. Congress passed the Hatch Act to establish Agricultural Experiment Stations to support research. In 1890, the Morrill Act was amended to establish land grant colleges to serve the educational needs of the African American population. In 1914, the U.S. Congress passed the Smith-Lever Act, designed to develop practical applications of research and providing instruction or demonstration of existing or improved practices (Rasmussen, 1989). The Smith-Lever Act led to the establishment of Cooperative Extension Services within all of the land grant universities. As a result, teaching, research, and extension became mandatory functions of land grant universities. Reflecting on the early history of land grant institutions, Bonnen (1998) stated, "The land-grant system of colleges did not spring into existence as a coherent idea or set of institutions in one decade or even one generation of leadership. The land-grant college evolved as an idea and then an institution and a national system over seven decades between 1850 and 1920 (p. 28)."

In Latin America and the Caribbean, extension services were institutionalized after World War II. Extension services in Asian countries were established after the 1950s, soon after their independence (Swanson & Claar, 1984; Naik, 1968). In most African nations, extension services started in the 1960s and 1970s (Ejeta, 2010). Extension services were seen as a way to promote agricultural growth and development in the rural sector. In most countries, existing agricultural ministries were reorganized to include an agricultural extension unit. Extension began to promote the use of modern inputs, such as new seed varieties, fertilizer, and pesticides, by training farmers, organizing method and result demonstrations, and making extensive use of the mass media.

Today, extension education exists throughout the world. Extension programs are generally the interface between local people and the government. Most government departments, such as agriculture, health, education and rural development, have organized extension programs to serve their clientele.

The function of extension is to enhance learning in non-formal educational settings. Knowledge can come from peers or the community, or from outside agents/experts. One function of agricultural extension is to communicate agricultural research findings and recommendations to farmers. Equally important is to communicate farmers' problems and needs to agricultural researchers and government policy-makers. Extension enables people to interact with each other and outsiders to gain information, insights, knowledge and skills to improve their capacity to solve problems, improve productivity, and upgrade their quality of life.

The extension programs designed for farmers are known as agricultural extension. Agricultural extension is part of the Cooperative Extension Service in the United States. In European countries, agricultural extension is called advisory services. In most developing countries, agricultural extension programs and services are organized under the respective ministries of agriculture. Over 150 countries have a national system for agricultural extension. Over 300,000 professional staff work for government extension services (Swanson, 1984) and more are engaged through non-governmental organizations and other donor-funded projects.

Agricultural extension is organized in many ways. Countries have set up different types of agricultural extension systems based on purpose, context and external support. Most agricultural extension services work in collaboration with agribusinesses, such as seed, fertilizer, pesticides, and production credit, to focus on technology transfer. Frequently, extension services emphasize advisory work by responding to requests from farmers and agribusiness operators. Often, extension services support human resource development and facilitate empowerment (Swanson & Rajalahti, 2010). In many instances, extension services offer all of the above four kinds of services to their clientele and/or stakeholders.

Because of multiple roles, functions and delivery mechanisms, people have difficulty understanding and defining agricultural extension. For the purpose of this training module, agricultural extension is defined to include all forms of non-formal education, mainly for rural people—farmers and agribusiness operators—aimed at improving farming methods and techniques, increasing production efficiency and income, and enhancing quality of life in a sustainable way. Its focus is primarily agricultural, i.e., about new and improved methods of crop and livestock production, processing, and marketing, and about fisheries, forestry, watershed management, and rural development.

Agricultural Extension and Advisory Service Models and Approaches

Agricultural development in developing countries has received support from many bilateral aid organizations, multilateral agencies, and foundations such as the United States Agency for International Development (USAID), the United Kingdom Department for International Development, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) of Germany, the Netherlands Development Organization (SNV), the Japan International Cooperation Agency (JICA), the United Nations Development Program (UNDP), the World Bank, and the Ford Foundation. Agricultural extension was and is believed to be an essential component of agricultural development. Based on donor preferences and recommendations, various extension models and approaches have been adopted.

Technology Transfer Model: Most extension systems rely on technology and information that are either available or able to be derived so as to be used by farmers. This form of extension relies heavily on a linear concept of technology transfer, i.e., new technology and knowledge generated by scientists/researchers/others is transmitted by extension agents to farmers to increase production

and income. This is the most common agricultural extension approach followed by developing countries.

Training and Visit Extension Model: Beginning in the late 1970s, the World Bank introduced the "training and visit" approach in about 70 countries to speed the dissemination of Green Revolution technologies to farmers (Swanson & Rajalahti, 2010). This approach assumed that extension educators were poorly trained and not up-to-date on the subject, poorly supervised, and tended not to regularly visit farmers. To address these problems, this approach introduced a system of regular training of extension staff by subject-matter specialists, regular visits by extension workers to innovative farmers in the community, and periodic interaction between farmers, extension workers and research scientists to facilitate two-way flow of communication.

Farmers Training Model: Agricultural extension programs in many countries initiated farmer training centers where select "model farmers" from surrounding villages or districts could get training about improved methods and techniques of farming. It was assumed that, after the training, the model farmers would go back to their villages, adopt the new farm practices they had learned during the training, and meet with others in the village to share what they had learned. These training programs would address two issues, 1) the inadequate number of frontline extension educators to serve a large number of farmers, and 2) the youth and inexperience of extension educators in the field. Farmer Field Schools are an adaptation of this approach.

Participatory Extension Models: There has been evidence that when rural people organize for their own benefit, much can be achieved. Generally, participatory extension approaches assume that local farmers have wisdom or indigenous knowledge regarding food and fiber production on their land, but their productivity and livelihood could be improved by learning more of what is known outside their locality or from applying scientific investigation techniques through on-farm trials (Axinn, 1988). Participatory Research and Extension makes the same assumption regarding the value of local knowledge and strives to create co-learning opportunities among extensionists, researchers and farmers (Percy, 2005). Most participatory extension models are supported by international NGOs and field activities are managed by local NGOs. Examples of this approach are Community Forestry, the Small Farmer Development Program, and the Farmers Associations of Japan.

Farmer-based Extension Organizations: In more developed economies, farmers' associations or cooperatives have established and managed agricultural extension programs to serve the needs of their members. Depending on the country, these extension systems operate under different management structures and with different sources of financial support. In general, members of the group or cooperative, not the government, control the functioning of the extension system. Members of the commodity group, such as coffee, sugar, cotton, or rubber growers, pay annual dues or a small portion of the amount of sale of the produce to receive extension services. In other situations, participants pay part of the cost of extension programs and government sources provide matching support.

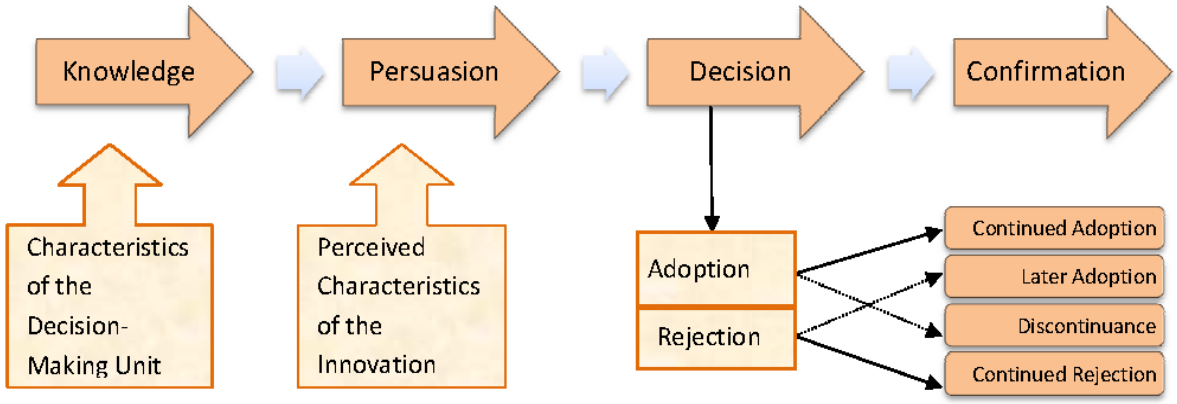
University-based Extension Model: Many agricultural colleges and universities offer outreach or extension services that help local communities, but also create an opportunity to improve the quality and relevance of their teaching and research functions. Agricultural universities have an assumed mandate to create and test technical knowledge so that it is relevant and useful to farm people. Also, both teachers and students benefit greatly from interaction with farmers. The U.S. land grant universities were developed to serve the people in each state with three interrelated and complementary functions: teaching, research, and extension. This concept is recognized by many agricultural colleges and universities developed through USAID assistance, and these universities have organized extension services in nearby communities.

It should be noted that a country could adopt a variety of approaches to agricultural extension to ensure technology transfer among farmers. Additional methods of human resources development, such as 4-H youth development, leadership training, consumer education, and formation of agricultural cooperatives, are also promoted by extension systems.

Agricultural extension has played an important role in development. Its role in testing and disseminating research-based agricultural knowledge and technology to rural people has resulted in improvements in the agricultural sector. Extension has facilitated the dissemination of messages about new varieties of crops, new breeds of livestock, and associated production and management practices, including use of fertilizer, implementation of irrigation systems, and marketing of farm products. Various communication channels and adult teaching methods have been used by extension staff to introduce new technologies to rural people.

The dominant paradigm of practice for extension workers has been the innovation decision process developed drawing from research studies in the fields of communication, anthropology, rural sociology, and extension education (Rogers, 2003). However, the innovation-decision process assumes that larger and wealthier farmers, who are innovators or early adopters, take advantage of extension services earlier than poorer farmers.

Figure 1: Innovation Decision Process



Adapted from Rogers, E.M. (2003)

With the ability to harness the insights of farmers and researchers to solve local problems, agricultural extension has contributed to increased production of food and fiber worldwide. Many countries have achieved food self-sufficiency since agricultural extension services were instituted to disseminate technologies resulting from national and international agricultural research systems.

Several extension models or systems did not meet technology transfer expectations. Some showed promising results in the short term, but failed to show lasting results. Others ceased their operations soon after the donor commitment for the project was over.

Although extension services are credited for the Green Revolution, which contributed to economic improvement, critics argue that environmental problems such as acid rain and pollution increased simultaneously. The technical package of the Green Revolution – use of irrigation systems and improved varieties that require fertilizer and pesticides – has resulted in negative impacts on the environment and human health.

Agricultural Extension in the 21st Century

Food and nutrition security in developing countries is a major thrust of the international development community. To address food and nutrition security, millions of small-scale farmers, foresters, fishermen, and agribusiness operators need educational support to increase their productivity and income. International development practitioners and policy makers have recognized that agricultural extension is a key pathway to achieving this goal.

Advancements in agricultural research, education, and communication have a direct impact on agricultural extension. Originally, most technologies, such as new crop varieties or breeds of livestock, were developed by public research institutions and were disseminated primarily by public extension services. This is changing very fast. Today, many technologies, such as hybrid seeds, pesticides, and information technologies, are being developed by international and national private sector firms. Private sector entrepreneurs have succeeded in providing communication services through Internet, radio and television. Non-governmental organizations and civil societies are creating education and training services for farmers and agribusinesses. Based on this scenario, we predict that extension services through the 21st century will emerge increasingly as partnerships between public and private sectors, including non-governmental organizations.

There is no single dominant agricultural extension system today. New approaches that integrate elements of many extension models are evolving constantly. In most countries, the central government provides an overall policy framework for extension, but a variety of actors (e.g., public organizations, civil societies, and private firms) provide a range of services to farmers and agribusiness operators. As a result, pluralistic extension systems are now common in many countries.

Examples of key elements embraced by contemporary agricultural extension services include:

Privatization: In the United Kingdom, public extension service has evolved over time into a private consulting business. In the Netherlands, farmers provide the majority of the cost of extension. Other forms of privatization include cost-recovery, outsourcing, and contracting out extension services. In Costa Rica, for example, the government "provides farmers with extension vouchers, which can be used for getting advice from private specialists"(Qamar, 2003, p. 24).

Pluralism: Contemporary extension services recognize the heterogeneity of the farming community and the need for diversity of extension service delivery systems. Multiple organizations, both public and non-public, deliver extension services. Examples include extension services delivered by local NGOs and private seed companies in Mali, Nepal, and Bangladesh.

Decentralization: Extension services are planned and implemented by district- or sub-district-level governments as in the Philippines, Tanzania, and Indonesia.

Client Participation: In the 'old school' models of extension services, diffusion was accomplished through absorption of information; in 'new school' models, diffusion is achieved through active participation of farmers in the learning process (Davidson, 2003). Extension programs and services are managed by membership of user groups such as coffee growers' associations and vegetable cooperatives.

Demand-driven agricultural extension is the buzzword used today, which means that extension responds to what farmers or clientele ask for to satisfy their educational and informational needs. The hope is that clientele will value the educational input so much that they will be willing to invest their own resources to receive the service. According to Chipeta (2006), service providers using this approach would be accountable to the users and the users would have free choice of service providers. Swanson (2011) argues that, to make these institutional changes, public extension systems must become more decentralized, farmer-led, and market-driven.

In summary, agricultural extension has been a donor-driven and project-focused effort. It has been a supply- and technology-driven system. Extension has operated with limited human and financial resources in areas having poor infrastructure, such as roads and communication networks. Extension has helped transfer Green Revolution technologies and practices to end users. However, to achieve food security, extension needs to play a bigger role in enhancing farmers' learning in non-formal educational settings. Technology transfer through education, communication, and provision of support services should be its primary function to improve rural livelihoods. No single delivery model can serve best in all situations. Experience suggests that the key elements of sustainable agricultural extension systems are participatory and pluralistic service, decentralized operation, and a market-driven system.

Challenges to Evaluation of Extension Programs and Services

Agricultural extension programs and services in most developing countries were initiated with external support. In the 1950s and 1960s, the international development community extended generous support to establish agricultural extension services to increase food production to end hunger and poverty. Because the need to support developmental programs is increasing and resources are limited, there is increased competition among agencies for funds. This has resulted in greater expectations for efficiency and accountability for organizations. Policy makers and funding agencies increasingly demand information on how program funds are used and what those programs produce. In this context, there is a growing need for program evaluation.

Program evaluation addresses this growing demand for accountability. Evaluation is a comparatively new discipline. Designs and methods for empirical studies are evolving constantly. Qualitative approaches and mixed methods are becoming increasingly necessary to assess and report effectiveness or document a program's impacts. This is particularly true in the field of extension education.

There is no uniform extension service delivery model. Some countries continue to follow the general agricultural extension service model – a top-down, technology-driven extension system. Others are adapting a decentralized (bottom-up), market-driven extension system. More recently, nongovernmental organizations and private-sector firms have entered extension work.

The roles and functions of extension services also differ by country and context. National food security is the priority for many developing countries. Accordingly, these countries expect extension services to disseminate yield-increasing technology to farmers. Most extension services focus on non-formal education for farmers. They utilize interpersonal and mass communication methods to aid in the dissemination of new agricultural information. Some extension services emphasize civic education through youth programs, leadership development, consumer education, natural resources management, or establishing farmer organizations to support rural livelihoods. Evaluation of these multiple roles and services is complex and requires multiple strategies.

Extension programs and services have multifaceted structures for funding, staffing and management. Extension performance is linked closely with that of agricultural research and input supply agencies, such as seed companies, fertilizer and pesticide dealers, agricultural credit banks, and marketing firms. In most cases, extension builds extensive partnerships and collaborations with government agencies, nongovernmental organizations (NGOs), other farm organizations, and cooperatives to plan and conduct programs. Also, the time required for adoption differs by technology or innovation, and support services such as government policy and the market for farm products. In this context, it is often difficult to attribute impacts of extension within the complex systems. GFRAS (2011) states the challenge as:

> The value of extension in stimulating the adoption and diffusion of new technologies is related to if and how extension has worked with research, the private sector, and farmer organizations to analyze and adapt new technologies to farmer needs and market demands (p. 24)."

The most challenging task for extension is to survive and thrive in a rapidly changing world. New technologies for production, processing, and marketing are introduced almost every year. Information and communication technologies are evolving constantly. The development and proliferation of social media have greatly impacted communication behavior of individuals and organizations. Extension services have lagged behind social media in the gathering, vetting, and rapid dissemination of information. Commodity prices change on almost a daily basis and are driven by international markets. Consumer preferences for food and beverages shift over time. Although yield increases have been realized and efficiencies have been enhanced, concern for the environment has increased simultaneously.

Evaluation of agricultural extension and rural development programs is a relatively new discipline. Despite the recognition that evaluation is important to build programs on solid ground, there has been no systematic effort to build capacity in this area. Many countries lack professionals who can plan and conduct systematic assessments of the results and impacts of extension programs. Agricultural extension programs have been utilizing donor-funded expatriates for program evaluation. These donor-driven evaluations, typically conducted toward the end of a project, often overlook local and indigenous perspectives. In many instances, national policy makers and program managers lack understanding of the theory and practice of program development and evaluation.

QUESTIONS FOR DISCUSSION

1. *What is the nature of agricultural extension or advisory services in your country? When did it start? Who funds this service? Is the service funded by one or by multiple sources?*
2. *What is the overall objective of extension service in your country? What model or approach does it follow?*
3. *How many professional extension staff work for the service in your country? What is their educational background and training? Does your extension service support staff development to upgrade their professional competencies? If yes, how?*
4. *What are the major challenges facing agricultural extension and advisory services in your country? How could these challenges be addressed?*

CHAPTER II

Introduction to Program Evaluation

Evaluation is not a new concept. It is something we all do, informally or formally. It involves thinking to make choices. It involves making value judgments.

We frequently engage in informal evaluation. Whether we are buying a computer, a television, or clothes, we evaluate options based on several criteria, such as our needs, size, and budget. We may read Consumer Reports magazine or a website to aid in decision-making. When we choose a career or a spouse, we use some other criteria to reach a decision. Sometimes we ask our friends or families for their opinions. In all of these scenarios, we are making judgments.

In a formal setting, we conduct evaluations by gathering information through a systematic process. We apply appropriate criteria or standards to this information to arrive at an informed judgment. We make the findings public to defend our conclusions.

The practice of evaluating the effectiveness of social programs in the United States began in the 1930s. Although evaluation of select social programs started in the 1940s and 1950s, assessment of The War on Poverty program initiated in the mid-1960s marked the beginning of large scale government-funded evaluation. The passage of the Elementary and Secondary Education Act of 1965 required schools or grant recipients to file an evaluation report (Weiss, 1998). About this time, evaluations of technology were conducted to determine the "bang for the buck" from investment through cost-benefit studies. Other economic impact assessment models followed.

Public interest in evaluation and accountability has grown steadily. Managers of foundations, civic organizations, and policy makers started raising questions such as:

"How is your extension program or project doing? What have you achieved?

"We have supported this extension project for five years; why should we continue this support?"

"What programs or activities of your agency have been effective? What problems have arisen? What are you doing to improve or terminate ineffective programs?"

"What new programs need to be developed to meet the needs of the people you intend to serve?"

Evaluation helps answer these questions. Evaluation is a management tool that measures and reports the results of programs and projects. It involves judging a program's merit or worth. It pinpoints the improvements needed by a program. Weiss (1998) defines evaluation as "the systematic assessment of the operation and/or the outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contribution to the improvement of the program or policy" (p. 4).

Case, Andrews, and Werner (1988) argue that "to evaluate is to make an explicit judgment about the worth of all or part of a program by collecting evidence to determine if acceptable standards have been met."

For the purpose of this guide, program evaluation is a continual and systematic process of assessing the value or potential value of extension programs or policies to guide decision-making for the program's future.

These definitions include three key elements of evaluation:

Evidence, or information: gathered systematically about a program or policy;

Standards, or criteria: used to judge a program or policy; and

Judgment, or assessment: reached when evidence is compared against standards.

WHEN WE EVALUATE . . .

- We examine the context and assumptions upon which a program or policy is based.
- We study the goals and objectives of the program or policy.
- We collect information about a program's inputs and outcomes.
- We compare it to pre-set standards. We make judgments about the program or policy.
- We report findings in a manner that facilitates their use.

Evaluation is both an art and a science. The art of evaluation involves identifying purposes and audiences, creating appropriate designs, and interpreting data about a program or policy. The science of evaluation involves systematically gathering and analyzing evidence about the outcomes and impacts.

Extension programs and policies are designed to reach certain goals and beneficiaries. Evaluations assess the programs and policies and help determine if goals have been achieved or benefits have been realized by the intended audience. Because agricultural extension has an educational delivery and technology transfer mission, evaluation in extension looks at changes in awareness, knowledge, skills and/or behaviors in targeted audiences, whether they are individuals, families, work groups, organizations, or communities.

Extension programs, no matter how large or small, need to be assessed to determine if they accomplished the stated objectives. Through evaluation processes, we find out what impact the program or policy had on the audience – how they reacted, what they learned, and whether the program was worth the time, money and resources invested.

First, let us differentiate between program evaluation and policy evaluation. A program typically includes subject matter focus, human resources, and infrastructure such as office space, staff, or vehicles. On the other hand, a "policy" is more likely to be a regulation or standard, with or without infrastructure. For example, a program to assist families with filing federal income taxes would need an office and staff, whereas a subsidy for chemical fertilizer could become a policy without the need for infrastructure.

The impact of an income tax assistance program could be measured by the number of tax returns filed, timeliness of filing, amount of tax collected, or number of citizens trained to file income tax returns correctly. On the other hand, the impact of the fertilizer subsidy could be determined by the increase in

amount of fertilizer used, increase in crop production due to fertilizer application, and/or level of export or import of food crops.

Why Evaluate a Program?

Agricultural extension programs are generally supported by public funds and/or through donor assistance. In recent years, the demands on extension for proof of program effectiveness and for public accountability have increased. Evaluation can help meet these demands in various ways. Also, organizational improvement cannot occur without evaluation. Only through evaluation can one learn: how well a program is doing, is there room for improvement, or in what direction should it be moving?

EVALUATION IS CONDUCTED FOR VARIOUS REASONS. THE PURPOSE(S) INCLUDE:

- Planning for change:
  - To assess needs and/or issues facing the community we are working with.
  - To set priorities to better direct allocation of resources.
  - To guide policy development.

- Analysis of program effectiveness or quality:
  - To determine achievement of project objectives.
  - To identify strengths and weaknesses of a program.
  - To determine if the needs of beneficiaries are being met.

- Effective decision-making:
  - To improve program management and effectiveness.
  - To expand or to terminate a program

.

- Maintaining accountability to stakeholders, funding sources, and the general public.

- Discover a program's impact on individuals, families, organizations and/or communities.

- Advocacy purposes:
  - To gain support from policy makers, advisory councils, and donor communities.
  - To direct attention to needs of particular stakeholder groups such as women farmers, small agribusiness operators, fishermen, etc.

What is the Role of the Evaluator?

The role of an evaluator is continually expanding. The traditional role of an evaluator was a combination of expert, scientist, and researcher who uncovered clear-cut cause-and-effect relationships. Today, evaluators are often educators, facilitators, consultants, interpreters, mediators and/or change agents. Evaluators could be internal or external to an organization

Patton (1997) asserts, "The evaluator is also a stakeholder – not the primary stakeholder – but in every evaluation, an evaluator's reputation, credibility, and beliefs are on the line (p. 364)."

As professionals, evaluators should take into consideration the needs and interests of primary users of the evaluation. S/he should coach an internal team through the steps of the evaluation as well as conduct certain parts of it. This will enhance credibility as well as use. Remember, when the concerns of primary stakeholders have been incorporated into the evaluation process, evaluation findings are more likely to be used.

AN EVALUATOR'S CREDIBILITY

An evaluator is judged by his or her competence and personal style. Competence is developed through training and experience. Personal style develops over time through a combination of training, experience and personal characteristics.

COMPETENCE

- Capacity to fully understand a program's context, goals and objectives
- Conceptual skills to design the evaluation
- Mastery of qualitative and quantitative approaches to data collection
- Basic quantitative and qualitative data analysis skills
- Report writing and presentation skills

PERSONAL STYLE

- Communication skills – verbal and written
- Confidence in the use of evaluation methods for project or policy being evaluated
- Strong interpersonal skills
- Ability to nurture trust and rapport
- Sensitivity in reporting evaluation findings
- Cross-cultural skills if engaged in international evaluation

What are Common Evaluation Approaches and Types?

Scholars differ greatly in their views of what evaluation is or is not, how it should be conducted, reported, and utilized. Some opt for a systematic approach to collecting information to assist decision makers. Others view evaluation as synonymous with professional judgment based on expert opinion.

Evaluators hold different philosophies about knowing or establishing truth, popularly known as epistemology. Epistemology is the theory of knowledge or the study of the nature of knowledge, philosophy of knowing, or establishing truth. Based on epistemology, evaluation approaches could be grouped under two categories – objectivism and subjectivism.

Worthen, Sanders and Fitzpatrick (1997) present some key elements of each epistemology. According to them, objectivism requires that evaluation information be "scientifically objective." In other words, use data-collection and analysis techniques that yield results reproducible and verifiable by others. This notion is derived from empiricism. Information or data collected are value-free, which means that the inquirer removes all subjective elements (values) from the situation, and only objective facts are presented. Evaluators are aware of experimental or control groups, subject characteristics, and structured observational protocols.

Subjectivism, on the other hand, bases its validity on "an appeal to experience rather than to scientific method. Knowledge is conceived as being largely tacit rather than explicit (Worthen, Sanders, & Fitzpatrick, 1997, p.65)." Evaluation procedures are "internalized," existing largely within the evaluator in ways that are not explicitly understood or reproducible by others. Information could be value-bound, which means that the inquirer is aware of the roles that values play in a given study. Therefore, it is important for the readers or users of the information to take into consideration both the inquirers' and participants' values and beliefs.

Over the years, evaluation scholars and practitioners have espoused different approaches based on purpose, time of evaluation, methodological backgrounds, and who conducts the evaluation. Some commonly described approaches to evaluation include the following:

Objective-oriented Evaluation Approach: Developed by Ralph W. Tyler in the 1930s to evaluate educational programs and projects, the basic assumption of this approach is that evaluation is a process of determining the extent to which the objectives of a program or project have been attained (Worthen, Sanders, & Fitzpatrick, 1997). Program evaluation involves the establishment of objectives in behavioral terms, finding situations in which achievement of objectives can be shown, collection of performance data, and comparison of performance data with stated objectives.

Management-oriented Evaluation Approach: Developed by Daniel Stufflebeam in the 1960s, this context, input, process and product (CIPP) approach views evaluation as "the process of delineating, obtaining, and providing useful information for judging decision alternatives" Worthen, Sanders and Fitzpatrick (1997, p. 98). The CIPP approach may help program managers to make four different kinds of decisions:

- Context evaluation, to serve planning decisions: What needs to be addressed by a program to help in defining the objectives?
- Input evaluation, to serve structuring decisions: What resources are available and what alternatives should be considered?
- Process evaluation, to serve implementing decisions: How well is the plan being implemented? What barriers threaten its success?

- Product evaluation, to serve recycling decision: What results were obtained? How well were the needs reduced? What should be done with the program after it has run its course? (p. 98)

Consumer-oriented Evaluation Approach: Espoused by Michael Scriven in the 1970s, the focus of this approach is on developing evaluative information such as rating scales for [educational] products for use by consumers in choosing among competing products or services.

Expertise-oriented Evaluation Approach: This is a widely used approach that involves direct application of professional expertise to judge the quality of whatever endeavor is evaluated. Accreditation and formal/informal review panels are examples of this approach.

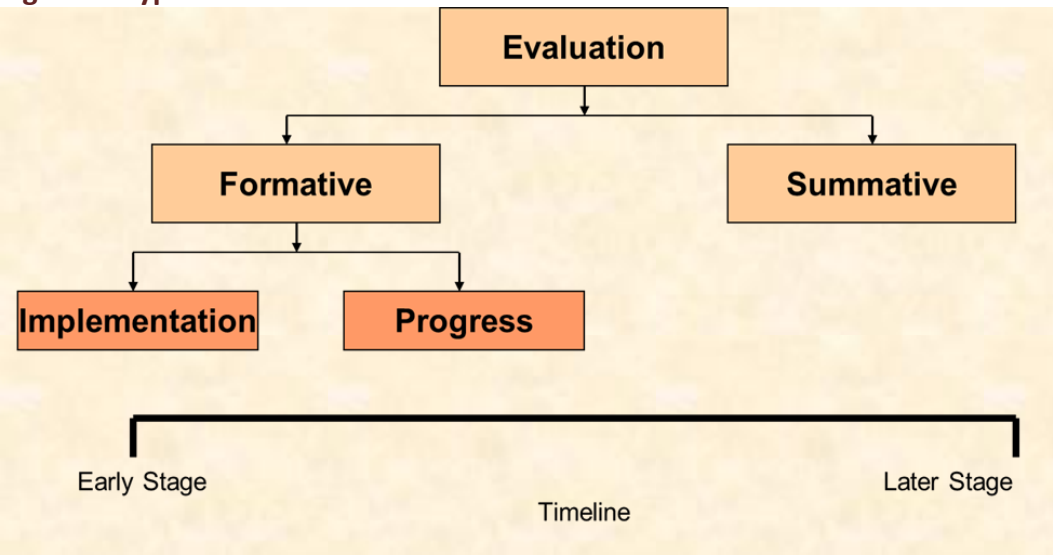
Participant-oriented Evaluation Approach: Developed in the 1970s, this approach considers involvement of participants as central in determining the values, criteria, needs, and data for the evaluation. The practice of stakeholder engagement in planning and evaluation, for example, resulted from this thinking.

The above approaches represent individuals' conceptions about the field of evaluation. They are neither models nor theories, and there is no one best approach to follow. Evaluation contexts and needs are different and it is difficult, if not impossible, to conceive of any one approach to be relevant to all situations. The choice of approach should be based on context, purpose, and resource availability, including expertise.

Based on general purpose, evaluations can be classified in two broad categories – formative and summative.

- A *formative* evaluation is conducted during the life of a program to identify its strengths or weaknesses and to enhance its quality and effectiveness.
- A *summative* evaluation is conducted at the end of a program to help decision makers decide a program's impact and its future. The focus is on determining program results and effectiveness (merit and worth). It serves the purpose of making major decisions about a program – continuation, expansion or reduction, and funding.

Figure 2: Types of Evaluation



Source: Frechtling, Frierson, Hood, & Hughes (2002)

Formative evaluation may take many forms. Usually these are evaluative studies conducted during the early stage of a program or project such as needs assessment, baseline studies, evaluation of on-going activities, or monitoring. Summative studies are usually planned during the later stages of a project (Figure 2).

Based on use of findings, Patton (1997) classifies evaluation into three categories: judgment-oriented, improvement-oriented, and knowledge-oriented. Impact assessments have judgment-orientation. Management generally values improvement-oriented studies. Donors and policy-makers seek knowledge-oriented evaluations and studies that answer accountability questions.

Based on timing and the specific purpose of the evaluation, evaluative studies could be grouped under three broad categories: evaluative studies conducted at the planning stage, during program implementation, and after the program ends.

EVALUATIVE STUDIES CONDUCTED AT THE PLANNING STAGE

Planners and policy makers conduct evaluative studies to gather input from various stakeholders of the extension program under consideration. These evaluative studies contribute to the development of program/project goals, objectives, strategies, and time lines. Results are used to develop feasible and realistic implementation plans. These studies also ensure that all stakeholders, including advisers, share a common vision of the program plan and of the evaluation plan. They help establish the groundwork for formative and later summative evaluations by developing measurable indicators and benchmarks.

The following types of evaluative studies often are conducted at the planning stage:

Context Evaluation: Context evaluation defines the bio-physical, socioeconomic, and cultural environment of the community in which the program/project will be presented. Its goal is to diagnose the needs or issues the program seeks to solve.

Needs Assessment: A needs assessment is an environmental monitoring process through which information is obtained that can be used to design timely, relevant, and reasonable programs (Fear 1988). It helps to establish which messages or media would work best to meet the identified need(s).

Input Evaluation: Input evaluations examine the human resources, budget, technology and equipment, facilities, and other resources that are necessary and available to deliver a program.

Feasibility Study/Market Analysis: The purpose of a feasibility study or market analysis is to determine if a program is feasible and/or desirable, and whether the available inputs and ideas can be crafted into a real-world program to benefit the intended audience. It also assesses how likely the program is to be successful in light of any other service providers who may be offering similar programs.

Benefit/cost analysis and rate of return on investment analysis are common methods used to determine the economic feasibility of new technology or development alternatives.

Baseline Study: Baseline studies measure the status quo, establishing a benchmark against which to judge future changes or program outcomes and impacts.

EVALUATIVE STUDIES CONDUCTED DURING PROGRAM IMPLEMENTATION

In-progress or Formative Evaluation: Formative evaluation looks at how a program is implemented and how the implementation process might be improved. This type of evaluation is conducted to make immediate changes or adjustments in the program. Formative evaluation usually takes place early in implementation and helps program managers find the strengths and weaknesses in a program while it is still going on. This type of evaluation is helpful for program improvement.

Mid-term or Midstream Evaluation: Mid-term evaluations are used when program managers want to adjust a program that is already underway. For a five-year project, mid-term evaluation is usually planned at the end of the second year so that adjustments in program design or delivery can be made to achieve project goals.

Monitoring: Monitoring involves gathering periodic information on project inputs and activities to ensure that the project is implemented as planned and enable management to take corrective actions when necessary. Program managers track resources (e.g., funds, personnel, and supplies) and processes (e.g., occurrence of meetings, demonstrations, and publications). For evaluation purposes, monitoring tracks key indicators of progress over the course of a program as a basis on which to evaluate outcomes of the intervention (Khandker, Koolwal, & Samad, 2010).

Misra (1998) offers 10 principles for monitoring:

1. ***Monitoring must be simple****.* A complex or complicated monitoring system is self-defeating.
2. ***Monitoring must be timely*** so that appropriate action may be taken.
3. ***Monitoring must be relevant*** to program objectives and generate useful information.
4. ***Information provided through monitoring should be dependable.*** Management will rely on monitoring findings only if the information is believed to be reasonably accurate.
5. ***Monitoring efforts should be participatory****.* It should include all concerned with extension, be they field-level personnel, subject-matter specialists, or extension's clients (the farmers).
6. ***Monitoring must be flexible****.* It is iterative in nature. It also becomes routine over time.
7. ***Monitoring should be action-oriented****.* It should follow pragmatic approaches, keeping the requirements of extension's clients uppermost in view.
8. ***Monitoring must be cost-effective****.*
9. ***Monitoring efforts should be top management-oriented****.* Monitoring units should keep in mind the requirements of top management when designing and operating a monitoring system.
10. ***Monitoring units represent specialized undertakings****.* Monitoring is not merely concerned with the collection and analysis of data, but with diagnosing problems and suggesting alternative practical solutions.

Operation Evaluation: Similar to monitoring, operation evaluation seeks to understand whether implementation of a program unfolded as planned. The aim is to compare what was planned with what was actually delivered to determine whether there are gaps between planned and realized outputs (Khandker, Koolwal, & Samad, 2010).

EVALUATIVE STUDIES CONDUCTED AT THE END OF THE PROGRAM OR AFTER

The purpose of program wrap-up, or summative evaluations, conducted at the end of or after a program, is to determine whether or not project objectives were met. These studies look for evidence of the value or success of a program. They measure the effects or impacts of a program. Summative evaluations also are called product, completion, or final evaluations. They supply unbiased information on the impacts, benefits, and cost-effectiveness of a program.

In some extension settings, this is the only evaluation conducted during the life cycle of an educational program. It summarizes what has occurred in the program, asks for end-of-program reactions, and attempts to assess success in meeting program objectives. It is used for program accountability purposes.

There are several different types of summative evaluation. Some of the commonly used include the following:

Output Evaluation: Output evaluation looks at basic program outputs, such as number of training programs conducted, extension bulletins published and distributed, number of male and female farmers reached, and program costs.

Outcome Evaluation: Outcome evaluation often measures progress in learning, such as changes in awareness, knowledge, attitudes, skills or behaviors. Usually, these studies focus on short-term impacts of a program, such as learning and medium-term impacts on people or policy. One example is development of a new policy to establish farmers' markets through cooperatives.

Impact Evaluation: Impact evaluation seeks to measure lasting impacts of programs or projects on important indicators such as crop yields, farm profitability, family income, or livelihood improvement. Such assessments also may focus on broad and long-term program effects, such as changes in ecological, social, economic, or community conditions.

Follow-up Study: A follow-up study is conducted long after a program is completed. This stage of evaluation looks at the long-term benefits of a program or policy. For example, participants in a leadership development program are contacted 5 or 10 years after completion of the training to determine whether and to what extent the training program was a factor in their career accomplishments.

QUALITATIVE VERSUS QUANTITATIVE STUDIES

Based on evaluation philosophy and methodology, evaluative studies can be qualitative or quantitative and follow ex ante or ex post facto methods.

Quantitative methods measure a finite number of predetermined outcomes and are appropriate for judging effects, attributing cause, comparing or ranking, classifying, and generalizing results. Quantitative methods are:

- suitable for large-scale projects;
- useful for judging cause and effect;
- accepted as more credible than qualitative methods by those who are oriented towards numbers; and
- applicable for generalizing to a larger population.

Quantitative methods commonly used in evaluation of extension programs include, but are not limited to:	
Existing information (e.g., census data)	Testing information & knowledge
Surveys (mail, telephone, online)	Economic models
Group-administered questionnaires	Personal survey/interviews

Qualitative methods take many forms, including rich descriptions of people, places, conversations, and behavior. The open-ended nature of qualitative methods allows the person being interviewed to answer questions from his or her own perspective. Qualitative methods are appropriate for:

- understanding the context in which a program takes place;
- understanding complex problems and process issues;
- clarifying relationships between program objectives and implementation;
- identifying unintended consequences of a program;
- gathering descriptive information;
- understanding operations and effects of programs; and
- conducting in-depth analyses of program impacts.

Qualitative methods commonly used in evaluation of extension programs include, but are not limited to:	
Existing information such as newspaper stories, family history, Focus group interviews	Key informant and semi-structured interviews Rapid rural appraisal
Participant observation	Case study

Mixed methods combine qualitative and quantitative methods within one evaluation study. This combination can be used to offset biases and complement each other by contributing the strengths of the varied methods. When using multiple methods, care should be taken to ensure that the selected methods are appropriate to the evaluation questions and that resources are not stretched too thinly. Multiple methods are appropriate for:

- understanding complex social phenomenon;
- allowing greater plurality of viewpoints and interests;
- enhancing understanding of both the typical and unusual case; and
- generating deeper and broader insights.

INTERNAL AND EXTERNAL EVALUATIONS

Based on who conducts the evaluation, either formative or summative, the evaluation could be conducted internally or externally to the organization offering a program.

- *Internal evaluations* are conducted by program/project employees. Generally, internal evaluations are used to monitor progress toward goals and for organizational learning.
- *External evaluations* are conducted by outsiders who are considered experts in the subject-matter. Generally, external evaluations offer more credibility than internal evaluations.

How do managers decide whether to use "in-house" staff to conduct evaluations or to hire external evaluation consultants? Table 1 below shows the advantages and disadvantages of using internal and external evaluators.

Table 1: Using Internal Versus External Evaluators

Options	Advantages	Disadvantages
Internal staff as evaluators	• Familiar with organization • Credible within organization • Develops institutional memory • Can follow up on evaluation recommendations	• Potential for lack of independence • Perceived organizational bias • Burden of additional tasks • Potential lack of power • May lack evaluation skills
External consultants as evaluators	• Specialized skills • New perspectives • Independence and objectivity • Readily available skills • Facilitation of program accountability	• Lack of knowledge of organization • Limited access to information and people • Lack of ability to follow up on recommendations • Can be expensive

Source: Boyle and LeMaire (1999)

When evaluators are evaluating their own programs, there are fewer problems involved in implementing findings. However, when evaluators are not the ones conducting the program, the likelihood of evaluation findings being ignored is greater.

One option is a "hybrid" – using internal staff and, as appropriate, contracting out portions of the evaluation to a professional evaluator. This hybrid approach helps organizations develop their internal evaluation capacity if external evaluators serve as coaches or facilitators.

EVALUATIVE STUDIES TO MEASURE LONG-TERM CHANGES

Agricultural extension and rural advisory services are long-term programs to serve farmers and agribusiness operators. As indicated earlier, many of these programs focus on education and technology transfer. Because learning and behavior change occur over time, evaluators generally are faced with the challenge of measuring long-term changes in participants or communities due to the program in place. The evaluators make use of longitudinal surveys to measure changes over time.

At a minimum, evaluators like to compare current conditions with past conditions using the same or similar measures and tools. Usually, they make the comparison with baseline data, census data, results of content analysis, or they track monitoring data from office records to assess change via select indicators. For example, they compare pre- and post-project data about awareness and knowledge of new farming practices, level of adoption or practice change, cost of production and yield, work-load for women members of household, marketing practices and costs, net profit, and improvement in livelihood.

Longitudinal Surveys: Longitudinal studies help track long-term changes. Three types of survey designs are popular: panel studies, trend analyses, and cohort studies.

> **Panel Studies:** In panel studies, the same subjects are surveyed at different times over an extended period. Although keeping track of people over time is difficult, data gathered from the same sample at different points in time offer meaningful conclusions.
>
> **Trend Analyses:** Different people from the same general population are surveyed at different times to detect if a trend is developing (e.g., adoption of IPM practices in vegetable production, adoption of hybrid corn, or membership in local food marketing cooperatives). Secondary data can show a trend, but cannot explain the cause. Evaluative studies that use a randomized experimental design with control group have the power to explain whether the trend is due to a new agricultural extension program or a new agricultural policy.
>
> **Cohort Studies:** These are longitudinal studies in which a specific population is followed over a period of time. Whereas trend studies sample a general population that changes in membership over time, a cohort study samples a specific population whose members do not change over the course of the study (Ary, Jacobs, & Razavieh, 2002). Cohort studies show if and how things have changed within a segment of the population over time.

Cross-sectional Studies: Cross-sectional studies are studies conducted at a single point in time from a specified population (Weiss, 1998). They measure a certain characteristic(s) or indicator(s), such as knowledge, attitude, or adoption behavior, using surveys, tests, or exams. For example, the evaluator may assess farmers' perceptions of hybrid maize in 2011, 2014 and 2017.

Pre- and Post-program Studies: Pre- and post-program assessment is the common approach to measure changes in specified variables as measured "before" and "after" a program takes place. A pre- and post-program study using a randomized experimental design with control group can explain whether the program has had any effect on the participants. Sample questions for pre-/post-program assessment are: Do farmers who attend extension meetings adopt hybrid corn varieties earlier than those who do not? Do they generate higher net income per hectare?

Does Evaluation Involve Research?

Evaluation professionals use a wide array of research methods, as diverse as casual observation and randomized experimental design. The word "data" refers not only to numbers, but also to content from

interviews, document reviews, observation, and case studies. Although knowledge about in-depth statistics is not often necessary, program managers should be able to identify cause-and-effect relationships between an activity designed to induce change (such as a workshop) and a particular desired outcome (such as increased knowledge of participants).

There are two schools of thought about evaluation of social programs such as agricultural extension. One school believes that evaluation involves value judgments and, thus, absolute accuracy is neither necessary nor attainable. Therefore, evaluation should be structured to serve as a learning process. Evaluators should be careful in the use of evaluation principles to improve judgments and decisions.

The other school considers that evaluation is useful only insofar as it provides credible evidence to inform real-world decision making. This requires (a) sound evaluation design, (b) valid and reliable information gathering about a program's outcomes and impacts, (c) credible methods to analyze and interpret data, and (d) conclusions derived from analysis of empirical data. To guide such evaluative studies, extension managers and policy makers must understand basic research methods and designs.

USAID puts much emphasis on measuring and documenting project achievements and shortcomings through evaluation. The new evaluation policy of USAID (2011) defines impact evaluation as that which:

> "... measures the change in development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a "treatment" or a "control" group provide the strongest evidence of a relationship between the intervention under study and the outcome measured" (p. 4).

Most evaluators consider evaluation as applied research. Research and evaluation both are grounded in empirical techniques. Both apply a systematic inquiry process. However, the intent of research and evaluation is different. The aim of research is the discovery of knowledge; evaluation in extension focuses on the impact of the application of knowledge on livelihoods and society as a management tool for decision making.

WHICH EXPERIMENTAL DESIGNS ARE USEFUL FOR EVALUATION?

What is an experiment? An experiment is a scientific investigation in which "the researcher manipulates one or more independent variables, controls any other relevant variables, and observes the effects of manipulations on the dependent variable(s)" (Ary, Jacobs, & Razavieh, 2002, p. 276). Remember two important terms:

- *Independent variables* are manipulated (changed) by the experimenter (e.g., the amount of nitrogen fertilizer applied kg/hectare to rice crop.)
- *Dependent variables* are observed but not manipulated by the researcher. They are variables upon which the effects of changes in independent variables are observed (e.g., rice yield per hectare).

In the context of extension evaluation, methods and strategies utilized by a program could be considered as independent variables whereas outcomes/impacts may be considered as dependent variables.

Also, there are three essential ingredients in the conduct of an experiment**.**

- *Control* of context and variables provides a situation in which the effect of a variable can be investigated.
- *Manipulation* is the process of setting up different treatment groups or conditions to facilitate observation of the impacts of independent variables on dependent variables.
- *Observations* are made with respect to specified characteristics/conditions.

When we make experimental comparisons, we usually begin with a hypothesis, which is a prediction that the experiment will have a certain effect. The experimental group receives a specific treatment while the control group receives no treatment. The use of a control group enables the researcher to discount many alternative explanations for the effect of the treatment.

Comparisons are essential in scientific investigation. Comparing a group receiving treatment with either an equivalent group receiving no treatment or an equivalent group receiving an alternative treatment makes it possible to draw well-founded conclusions.

Experimental design refers to the conceptual framework within which the experiment is conducted. It is the researcher's plan for carrying out the experiment. The experimental design serves two functions.

- It establishes the conditions for the comparisons required to test the hypotheses.
- It enables statistical analysis of data.

An example of an experimental design for evaluation is given in Figure 3.

Figure 3: Diagram of Experimental Design for Evaluation

Subjects Randomly Assigned?	Results of Pre-project or Pre-test Measurement	Treatment Group	Results of After-project or Post-test Measurement
Yes	A0	Project Participants	A1 [or Experimental group]
Yes	B0	Non-Participants	B1 [or Control group]

- A true experiment requires random assignment of subjects to a treatment group. Random assignment is the only way that groups can be considered statistically equivalent.
- In the above experimental design: A1-A0=Y (outcome due to treatment) and B1-B0=Z (outcome without treatment)
- If Y is greater than Z, the program has had a positive impact on net outcome. If Y=Z, the project had no impact on outcome. If Y is less than Z, the project had a negative impact on the outcome.

In agricultural extension and rural advisory service settings, data collection following a randomized experimental design is often not practical. Data collectors would face political and ethical issues in excluding some households within the same service area. As a result, the evaluators cannot easily observe the outcomes of a program on participants if they had not been the beneficiaries. Without information on the counterfactual (i.e., what would have happened to the participants of a program had they not participated), the best alternative is to compare outcomes of project households with those of a comparison group that has not been the beneficiary of the project.

Khandker, Koolwal, & Samad (2010) offer different strategies to address the counterfactual issue, including propensity score matching, double-difference method, instrumental variable method, regression discontinuity and pipeline methods.

While evaluating developmental programs, including extension education, it is frequently not possible to conduct a true experiment with random assignment of subjects. We often have to deal with intact groups such as 4-H clubs, neighborhoods, districts, or training groups. Designs that do not include random assignment are known as quasi-experimental designs. These designs permit us to reach reasonable conclusions even though full control is not possible.

Figure 4: Example of Quasi-Experimental Design or Non-Equivalent Control Group Design

	Before Project		After Project	Impact
Group A	A0	Treatment/ Project	A1	A1-A0=Y
Group B	B0	No treatment/ Control	B1	B1-B0=Z

- No random assignment **of subjects;** treatment **given** to Group A, but **not to the other group (**Group B**).**
- If A1 is greater than A0, the project has a positive impact on participants.
- If Y is greater than Z, the program had a positive impact on outcome.

Results of quasi-experimental design provide reasonably good evidence of program impacts.

Generally, program and policy evaluators in extension and rural development settings have not used experimental research designs. Pound, Gundel, Martin and Apenteng (2010) conducted a meta-evaluation of 17 national/regional extension evaluation case studies. The findings indicated that most evaluation studies used some form of questionnaire survey, review of relevant documents, key informant interviews, or case studies. Only one evaluation used random sampling of survey respondents.

USAID evaluation policy strongly recommends identification of key evaluation questions at the outset of a project to improve project design and guide data collection. The policy supports the collection of baseline data to establish a reference point. Baseline studies should collect sex-disaggregated data using household or individual surveys. And can be replicated toward the conclusion of implementation to assess change (USAID, 2011). The USAID policy emphasizes:

- Evaluation be integrated into design of projects
- Evaluation be unbiased in measurement and reporting
- Evaluation be relevant, i.e., consultation with partners and beneficiaries is essential
- Evaluation be based on best methods, i.e., consider empirical strength and feasibility
- Evaluation be oriented toward reinforcing local capacity
- Evaluation be transparent and findings be shared as widely as possible

Alternatively, evaluators may collect only "after-project" data from both groups (i.e., participants and non-participants) using a random sample from each group.

Figure 5: Example of Evaluation Design for a Training Program

Sample	Group	Score Before	Score After	Net Change
Random Assignment of Trainees	Experimental Group or Program Participants	P1	P2	P2 - P1 = P'
	Control Group or Non-participants	C1	C2	C2 - C1 = C'

Adapted from Weiss (1998)

In the above example, if P' is greater that C', the program has had a positive net outcome.

We believe that evaluation in extension and rural advisory service should be a learning process. Evaluation results should be utilized in improving the program and developing more effective policies to serve rural populations. Extension managers, therefore, should plan evaluation at the time of planning the project itself. The monitoring process should be developed to track key indicators of progress over the course of the project.

For practical reasons, we believe that not all agricultural extension programs and policies need experimental (or randomized) setups (and it is not feasible cost-wise to use it every time) to accurately document the impacts of the programs or projects on families, communities, and society. Alternatively, qualitative studies could offer the depth of information needed about the impact of programs or policies on the intended audience. A mixed method approach might offer better documentation of outcomes and impacts.

Assessing the feasibility of a program evaluation helps ensure that the program can be evaluated meaningfully and that the evaluation will contribute to improving program design and/or performance.

USAID's evaluation policy suggests that each operating unit will conduct at least one performance evaluation of each large project it implements. Evaluations are expected to make use of sound social science methods and include the following features (USAID 2011):

- Evaluation team with appropriate methodological and subject matter expertise
- Use of a written design including key questions, methods and dissemination plan
- Gender-sensitive indicators and sex-disaggregated data
- Participation by national counterparts and evaluators
- Use of reliable data collection and analytic methods
- Report findings based on facts, evidence and data
- Evaluation reports that include action-oriented, practical and specific recommendations

So, the art and craft of evaluation is fairly delicate. As a program evaluator, you have to juggle many factors and walk a fine line between management and politics.

Ethics of Evaluation

Evaluation data collection, presentation of results or findings, and management of information should follow ethical standards. Major considerations include:

Cultural Sensitivity: We should be aware of our own attitudes, beliefs and values to avoid or minimize biases. Evaluation protocols should be developed only after careful review of cultural norms and values. Recognize cultural differences.

Political Nature of Evaluation: We need to understand how evaluation is influenced by internal and external politics. How do evaluators maintain neutrality in the political atmosphere and to whom they are accountable.

Data Collection and Analysis: We have to address the ethical and privacy concerns of the people from whom we gather data. Always seek permission or informed consent from the participants. Obtain parental consent and participant assent if data are to be collected from minors. Do not collect information unless you will be using it. Obtain institutional approval for evaluative studies involving human subjects. Appropriate analysis is critical. Present information in a way that is easy to understand and free of jargon. Always protect and respect the privacy of participants and respondents.

Presentation of Findings: We should understand that not all information is useful to all stakeholders. Also, not all stakeholders are information users. So, getting the right information to the right people is essential so that information intended for specific uses is likely to hit the target. At the same time, telling the truth to people who may not want to hear it is another purpose of evaluation (Patton, 1997).

Safe Storage of Evaluation Data: Information collected for evaluation, whether quantitative in the form of surveys or qualitative in the form of recordings or pictures, should be maintained in a safe and secure storage place. The information should be disposed of appropriately after the program ceases or the information is no longer deemed useful for future follow-ups.

Popular Program Evaluation Frameworks and Designs

PROGRAM LOGIC MODEL

Below is an illustration that shows the major components of a program or project connected to the outcomes that are caused by or attributable to the program. Program logic is not a theory or an evaluation model; rather, it is a framework for describing the relationships between investments, activities, and results (Taylor-Powell, & Henert, 2008).

Logic models help us plan, implement, evaluate, and communicate more effectively. International aid organizations use logic models in developing new projects.

Logic models consist of a logical chain of connections showing what the program intends to accomplish with the help of a series of “if ___ then ___” relationships.

A logic model helps to clarify logical links between project inputs and objectives, project activities and outputs, broader purposes, and the ultimate goal. It is particularly useful for planning activities, resources, and inputs required to meet objectives (Suvedi, den Biggelaar, & Morford, 2003). The

program theory of action illustrates the link between what we do with the resources we have and the impacts of our activities. Figure 1 illustrates the program theory of action for a pesticide safety program. Note that logic models could include a narrative that explains the relationships between these components.

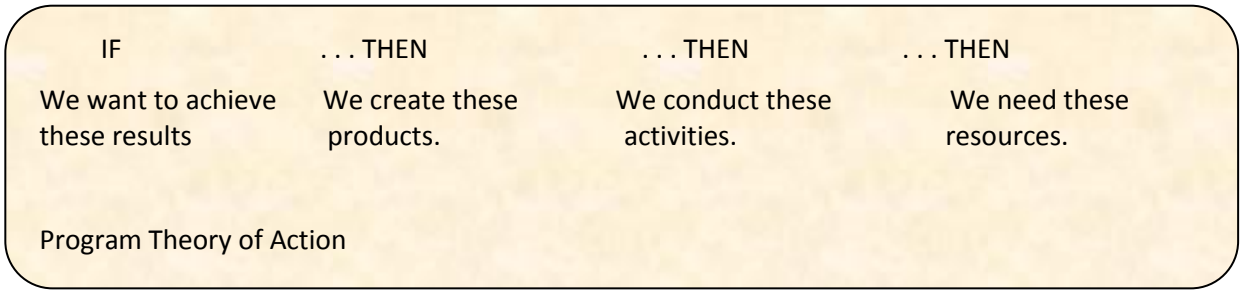
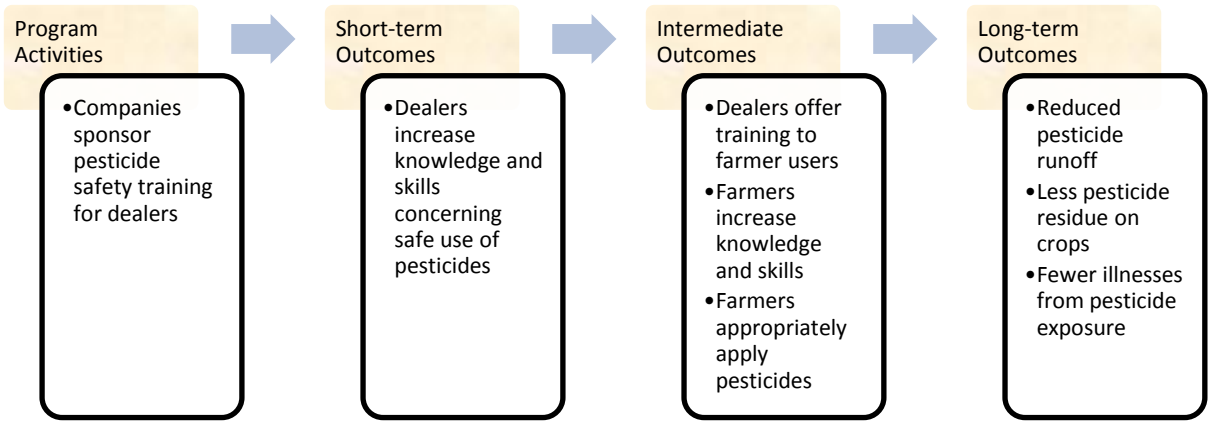


Figure 6: Illustration of a Program Logic Model



Key elements of a logic model**:**

Inputs: Inputs are the human, financial and material resources required to implement a program or policy.

Outputs: Outputs are the quantity of products and services delivered by the program or project to its clientele, such as number and types of programs, number of participants, memberships acquired satisfaction with program, and services.

Outcomes: Outcomes are the measurable results or consequences – both expected and unexpected – of an activity or program in meeting its stated goals and objectives, such as the percentage of participants who gain knowledge or skill as a result of the program.

Impacts: The fundamental intended or unintended change occurring in organizations, communities, or systems as a result of program activities. These could be short-term (e.g., awareness of learning), medium-term (e.g., adoption of a new practice or taking social action such as formation of a farm cooperative), or long-term (e.g., quality of livelihood, adoption of environmental-friendly farm policy).

We should note that an outcome at a lower level of impact may serve as an input for a program targeting higher level impacts. Figure 7 below depicts an outcome chain as an example.

Figure 7: Outcome Chain Example for Sustaining Rural Population

Benefits of Logic Modeling

- It helps to bring clarity to fuzzy goals.
- It shows a logical chain of events linking inputs and activities to outcomes.
- The key elements of program are summarized.
- The differences between activities and outcomes can be clarified and communicated.
- Evaluation questions can be identified easily.

Limits of Logic Modeling

- A logic model is a *representation* of reality, not reality.
- Programs are not linear and it can be difficult to graphically represent complex programs.
- Challenges of causal attribution – many factors influence outcomes.
- Doesn't address critical question of "are we doing the right thing?"

HIERARCHY OF EVIDENCE (HIERARCHY OF PROGRAM EVALUATION)

Bennett's Hierarchy of Evidence provides a way of conceptualizing the relationships between program objectives and outcomes at different program levels. The hierarchy shown below (Figure 3) suggests the kind of information appropriate to determine if an objective has been met. The "lowest" level of

evaluation (Inputs) refers to resources expended for programs or projects. At this level, evaluative studies report the number of person-days of staff time devoted to a project, money spent, or vehicles procured. This statement does not say very much about the difference that the program or project has made on the target audience. Ideally, evaluations want to report from a higher level on the ladder.

Further up the ladder, you document changes in knowledge, attitudes, skills, and aspirations (or plans) of the target audience. You might report about knowledge change such as the understanding impacts of IPM on vegetable production, based on pre- and post-test results. This kind of evidence is slightly more convincing to someone judging whether or not your program was worthwhile.

Further up the ladder, you can report on changes in practices or behaviors of your target audience. For example, "About three out of five (60%) farmers who participated in IPM training reported using the recommended practices within 2 years." This is even better evidence of a program's worth, but requires you to survey participants to determine if they incorporated the new practices into their operations.

The top rung of the ladder shows results related to the long-term benefits or impacts that drive the program or project – for example, "environmentally friendly agricultural production." Managers typically cannot evaluate at the top level because they cannot isolate other factors that may have led to the long-term result, but it is nevertheless helpful to know what the ultimate expected outcome or impact is.

Figure 8: Evaluating Program Performance

Adapted from Bennett (1979) and Bennett & Rockwell (1995)

Every program or project manager should establish desired outcomes at the outset of a program, project, or policy. The higher up the ladder, the more time and resources it takes to gather data about outcomes, but the more convincing the evidence will be. As a manager, you must decide the trade-off between strong evidence of worth and the cost/time required to gather evidence (Morford, & Suvedi, 2002).

The higher the level of hierarchy of evaluation, the more complex it becomes, and the need for expertise and resources to conduct the evaluation increases. The illustration below shows appropriate types of information to gather for the level of the program you are evaluating.

Table 2: Commonly Used Indicators for Program Evaluation

PROGRAM LEVELS	COMMONLY USED INDICATORS
SEEC Change or End Results	Changes in participants' personal and working lives as a result of program participation
Practice and Behavior Changes	Changes in participants' practices as a result of program participation
Knowledge, Attitude, Skill and Aspirational (KOSA) changes	Changes in participants' knowledge, opinions, skills and aspirations as a result of program participation
Reactions	Reactions of participants and clients to the program
Participation	Number and profile of program participants (e.g., male or females; large or small farmers, etc.)
Activities	Activities in which participants were engaged through the program. The kinds of information and methods used to interact with program participants
Inputs	The personnel and other resources used during the program.

Adapted from: Bennett, C. 1979. *Analyzing impacts of extension programs.* Washington, D.C.: U.S. Department of Agriculture, Science & Education Administration (ES C-575).

Rockwell and Bennett (1995) have developed an integrated model for program planning and evaluation. It focuses on outcomes in planning, implementing, and evaluating programs. The model is called Targeting Outcomes of Program (TOP) and is available online at: http://citnews.unl.edu/TOP/english/.

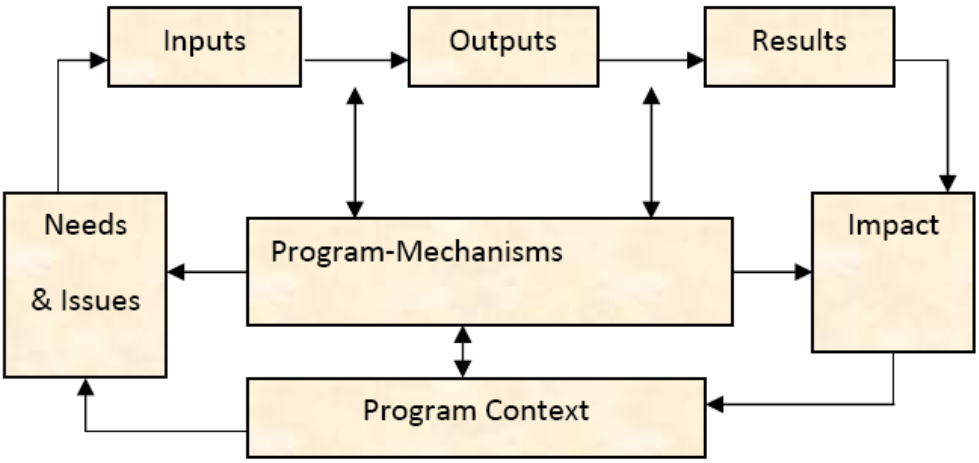
Program Evaluation Cycle

Evaluation is the process of determining the worth, or value, of a program or project. On the surface, the program logic and hierarchy of evidence models portray evaluation as a linear process. Actually, extension program development and evaluation is a reflective process involving many steps in a cycle.

At the international conference on Systemic Approaches in Evaluation, Hummelbrunner (2011) outlined the "systemic loop" to illustrate and guide reflective practice. He argues that the reflective practice involves actions based on prior hypotheses, and monitoring their effects enables adjustment of hypotheses and subsequent actions [with a view to what has been or remains to be achieved].

Hummelbrunner (2011) proposes a "circular" logic model by adding two components (i.e., program context and program activities/intervention mechanism) that are interrelated with the elements of the logic model. He argues that such a circular model can provide a framework for evaluations that are both theory-based and systemic. Ideally, evaluation can begin at any point in the cycle. Evaluation offers feedback with learning taking place through each repetition.

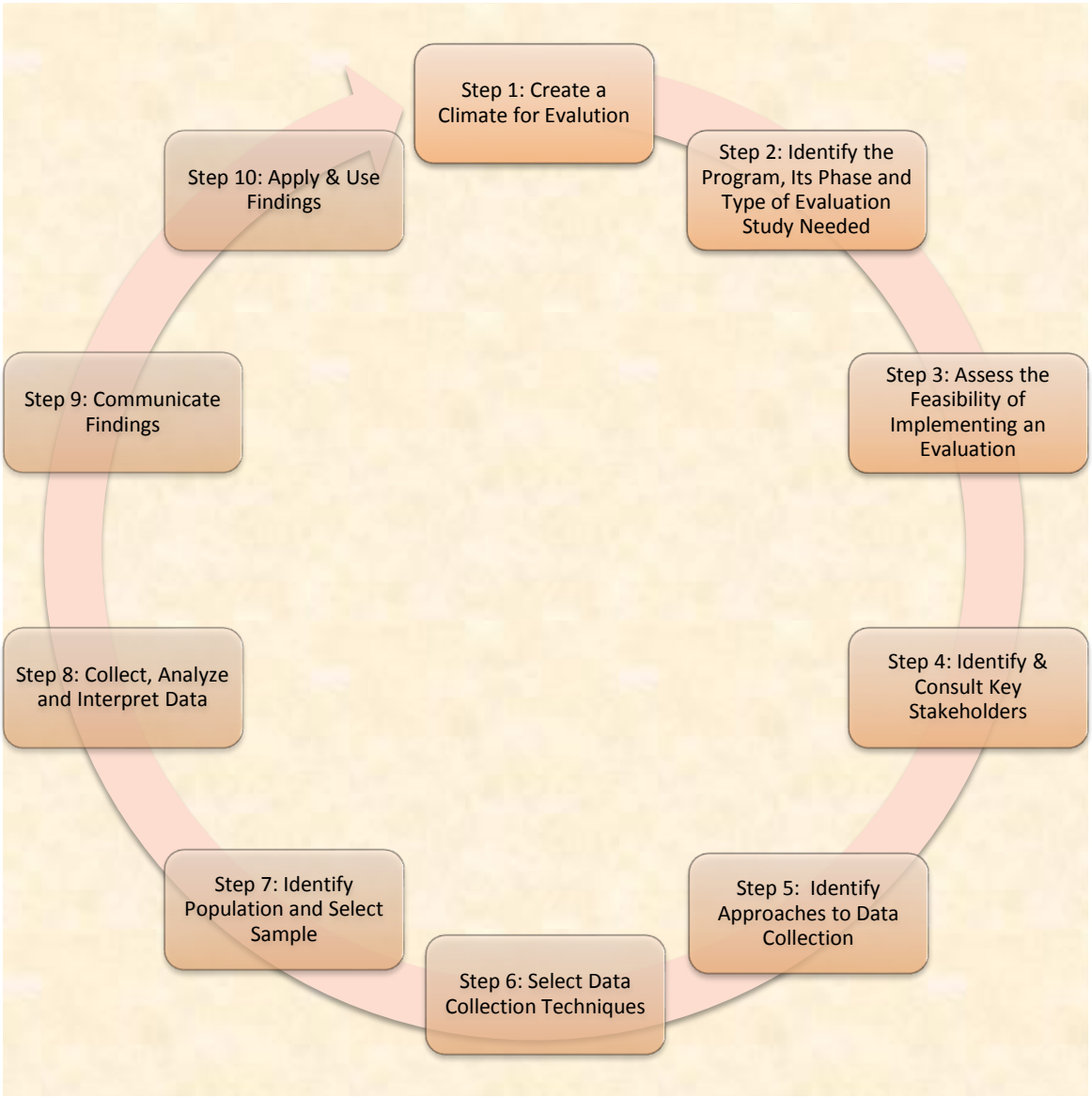
Figure 9: Circular Framework for Program Evaluation



Source: Hummelbrunner, 2011, p. 13

Evaluation practitioners have proposed a number of steps in the evaluation cycle. To make program evaluation less intimidating and more manageable, the cycle can be segmented into 10 manageable steps. However, note that the specifics of each step may vary, depending on the nature, scope and complexity of the program and the resources available for conducting the evaluation. Figure 4 shows a 10-step cycle for program evaluation.

Figure 10: Ten Step Cycle for Program Evaluation



The remainder of this training manual is organized by the steps identified in Figure 4.

Chapter III describes creating a climate for evaluation (Step 1). Chapter V presents steps in developing an evaluation plan (Steps 2 through 5). Chapter VI describes how to plan and manage data collection (Steps 6 through 8) and Chapter VII offers insights on how to analyze data, communicate results and use findings to improve programs (Steps 9 and10).

RESULTS FRAMEWORK

The "results framework" is used widely as a strategic planning and management tool by USAID in the broader context of programs and national goals. It is a graphical presentation of the expected results of a program or project. It includes objectives and the intermediate results of a program. Program managers can use the framework to identify the indicators by which a project's progress is monitored and evaluated, and the conditions necessary for the project to achieve the expected results. It is particularly useful for realistic planning of the activities, resources required to meet project objectives (Uribe, & Horton, 1993). It helps clarify the logical links between project inputs and objectives, project activities, outputs, broader purposes, and the ultimate goal. It conveys the development hypothesis implicit in the strategy (USAID 2000). The framework itself is not a plan for evaluation, but it is useful in designing program evaluation.

QUESTIONS FOR DISCUSSION

1. *What is your definition of program evaluation? Write down your definition on a piece of paper and share with your fellow participants.*
2. *If you were asked to evaluate an agricultural extension project, when would you start planning the evaluation? Why?*
3. *If you were to select a professional program evaluator, what qualifications or competencies would you consider important? List these competencies on a piece of paper and share them with your group.*
4. *If you were to conduct an evaluation of an agricultural extension project, what approach or type of evaluation would you follow? Why?*
5. *Does your organization support monitoring of extension programs? Are monitoring data used for program improvement?*
6. *Many country extension systems lack monitoring and evaluation systems apart from those required by donors. How can monitoring and evaluation systems be established in these countries?*
7. *Do you support internal or external evaluations for your extension program? Why?*
8. *How is evaluation different from research?*
9. *How practical is experimental design for evaluation of extension programs? If you were to plan an evaluation, what kind of design would you use? And why?*
10. *What evaluation model or framework best fits your context?*
11. *What are the major ethical dilemmas facing program evaluators in your organization? How can they be addressed?*

CHAPTER III

STEP 1: CREATING A CLIMATE FOR EVALUATION

Extension managers and educators have a primary role to play in the evaluation of educational programs. Each time we plan a program, we identify the program to be conducted, write objectives we want to accomplish through the program, decide on a plan of action to meet the objectives, and develop an evaluation plan to determine if the objectives were accomplished. A meaningful plan for evaluation outlines clearly the standards for evaluation or the criteria on which the program is to be judged. It also provides guidelines for what information should be gathered about the program, and how, when, where, and by whom it should be gathered.

As program managers, we should be knowledgeable of the basic principles and processes of evaluation so that we can empower our staff and educators to perform many evaluative functions for our educational programs. All educators should be able to plan and conduct their own evaluations. If we believe in the participatory mode of educational programming, we should function more like a midwife than as a mother. We need to guide others in the process and utilize the results of these internal evaluations for program improvements.

For various reasons, evaluation may not be a standard practice in some organizations or projects. Evaluation may also be intimidating. What if it turns out that existing programs aren't working? Will my project lose support if evaluations show poor results? Will the employees lose their jobs? These fears, often based on misperceptions, can become barriers and often create a climate that discourages evaluation.

Barriers to Evaluation

- Some organizations are reluctant to accept evaluation as an integral part of day-to-day work.
- Evaluation is feared as a threat to programs and program staff. Either intentionally or unintentionally, findings can be misused for political advantage.
- Disagreements occur within the evaluation process, i.e., what indicators to use, which designs to follow, and what type of evidence or data to gather.
- Problems occur in the sharing of evaluation results.
- Evaluation is seen as another thankless task by overworked employees.

Adapted from Marynoski, Denny, Colverson, & Hill (2006), page 23.

If an organization is not accustomed to evaluating programs, there is a need to build awareness and understanding of, and support for, evaluation. First, program managers must be convinced that evaluation helps to build successful extension programs. Second, management needs sufficient

motivation to commit the time and resources necessary to perform evaluations. Third, program staff must establish and communicate a unified evaluation purpose, to include one or more of the following:

- organizational and program accountability;
- program development and improvement;
- informing management and policy decisions; and/or
- organizational and staff learning.

Features of an Evaluation Culture in an Organization

Several key principles can be used to help create an evaluation culture within an organization.

- Programs that get measured get attention and action.
- Program successes must be measured to learn from them or unintentionally, findings can be misused for political advantage.
- Program problems must be measured to correct them.
- Program results must be demonstrated to win public support.
- Without measures of program results, you can't improve the efficiency, effectiveness, capacity, or quality of programs or organizations.
- Program results must be measured to distinguish success from failure. Otherwise you might unknowingly be continuing to reward failed programs.

Adapted from Marynoski, Denny, Colverson, & Hill (2006), page 16.

Successful program managers and administrators address barriers in a systematic way by building an evaluation culture within an extension organization. They start building evaluation support by engaging administrators, staff, and stakeholders in evaluation and rewarding their work. Staff members should receive support to attend evaluation training programs. Investing in staff evaluation skills is one of the best long-term investments an organization can make in program improvement.

Successful program managers link evaluation to organizational learning and performance review. Extension educators are encouraged to conduct evaluations so they know what they do has value and that they can always improve performance by reflecting on their experiences.

QUESTIONS FOR DISCUSSION

1. *Who performs the monitoring and evaluation function within your extension organization? What kind of monitoring and evaluation service does this unit or person provide? If there is no such unit or persons within your organization, do you see a need for such a unit?*
2. *Does management of your extension system support program evaluation? How could this administrative support be strengthened?*
3. *What are the major barriers to effective monitoring and evaluation services within your organization? How could these barriers be overcome or minimized?*
4. *How could donors and professional evaluators assist in strengthening the evaluation function within your organization?*

CHAPTER IV

Developing an Evaluation Plan

Planning an evaluation begins with identification of the program to be evaluated. Once you accept the evaluation responsibility, you should meet with program managers to get specific information about the program needing evaluation, including:

- goals and objectives;
- geographic boundaries of the program;
- clientele served;
- program funders; and
- program staff.

Broadly, there are three types of evaluative studies by program phase. Some evaluative studies are appropriate for the planning phase. Others are suitable for monitoring of progress during the implementation stage. Still others are used to document results and impacts towards the final years of a project. The type of evaluation study utilized is selected on the basis of stage of program, program requirements, and stakeholders' interests.

STEP 2: IDENTIFYING THE PROGRAM PHASE AND EVALUATION TYPE

Find out the stage of the program or project to be evaluated. Is it at the planning stage? Is this an on-going program? Is the program at the concluding stage? This information is helpful in deciding the type of evaluation.

Figure 11: Evaluative Studies by Program Phase

Ask:		**Identify program phase:**		**Select type of evaluation study:**
Is the program in the design stage?	↔	Planning phase	↔	Needs assessment
Is the program just beginning?	↔	Implementation phase	↔	Baseline study
Is the program active?	↔	Implementation phase	↔	Formative evaluation
Is the program ending?	↔	Final phase	↔	Summative evaluation
Is the program over?	↔	Final phase	↔	Follow-up study

Ask the program managers: "Who will be using the information from the evaluation"? Find out the current program phase and determine the appropriate type of evaluation study.

After we determine the stage of the program to be evaluated, we should choose an appropriate evaluation design based on answers to several questions. Will this be a one-time study, a pre-post study, or a longitudinal study for which data gathered in a series of studies? Are we interested in evaluation using experimental design or is a quasi-experimental design sufficient? Some prefer a case study to provide in-depth information about benefits of a program to an individual, family, or community. Others like to conduct focus group interviews with a cross-sectional sample of stakeholders.

STEP 3: ASSESS THE FEASIBILITY OF IMPLEMENTING AN EVALUATION STUDY

Assessing the feasibility of a program evaluation helps ensure that the program can be meaningfully evaluated and that the evaluation will contribute to improving program design and/or performance. Consider the following questions carefully and then decide whether this is an appropriate time to begin a program evaluation. If the answers to many of the following questions are "No," this may not be an appropriate time to implement an evaluation study:

- Is there an important decision to be made on the basis of the evaluation?
- Is there a commitment to use the evaluation findings?
- Will important program decisions be made regardless of evaluation findings?
- Is there a legal requirement to carry out an evaluation?
- Does the program have enough impact or importance to warrant formal evaluation?
    - Is this a one-time program?
    - Will this program continue?
    - Is the cost of the program so low that an evaluation is unnecessary?
- Is it likely that the evaluation will provide valid and reliable information?
- Is it likely that the evaluation will meet acceptable standards of propriety?
    - Will the evaluation violate professional principles?
    - Is the evaluation threatened by conflicts of interest?
    - Will the evaluation jeopardize the well-being of program participants?
- Is the program ready to be evaluated?
    - If a summative evaluation is suggested, has the program been operating long enough to provide clearly defined outcomes?
- Are there sufficient human and monetary resources available to carry out an evaluation?
- Is there enough time to complete the evaluation?

STEP 4: IDENTIFY AND CONSULT KEY STAKEHOLDERS

Evaluation stakeholders are people who have a stake or vested interest in the evaluation findings. They can be program funders, staff, administration, clients or program participants. They may not be the same individuals as the program stakeholders. It is important to clarify the purpose and procedures of an evaluation with key evaluation stakeholders before beginning. This process can help determine the type of evaluation needed and point to additional reasons for evaluation that may prove even more productive than those originally suggested.

Come to agreement with stakeholders about:

- which program will be evaluated, what it includes and excludes;
- the purpose of the evaluation;
- the goals and objectives of the program (goals and objectives can be written as statements indicating what the program will achieve and what criteria will be used to judge whether the objectives have been met; for example, see the box below);
- the budget and time available for evaluation;
- the evaluator's role; and
- who will receive the evaluation results.

Because objectives provide such a critical set of guidelines for evaluation, they should be stated clearly and succinctly. When program objectives are stated clearly, the indicators and criteria to judge merit or worth are stated explicitly. Objectives should indicate questions and issues the evaluation will address. They also indicate who will participate in the evaluation.

A good program objective should:

- contain one outcome;
- identify the target audience;
- specify what you expect to change as a result of program participation;
- be specific enough to be measurable and indicate how the objective will be measured

Example: *An extension project in District A aims to increase local rice production. The indicators will be that 75% of farmers will adopt a new variety of rice and apply chemical fertilizer within 3 years, increasing yield by 50%.*

Clarify Evaluation Issues, Questions, Indicators, and Criteria

Evaluations are conducted to answer specific questions or test hypotheses to judge the value, or worth, of an existing program. If the questions and issues being used are not clearly defined and the indicators and criteria that will be used to judge merit or worth are not well thought out, the evaluation may lack focus, be irrelevant, omit important areas of interest, or result in unsupported conclusions.

Basic Steps in Selecting Indicators and Criteria

- List questions, issues and criteria from all sources consulted.
- Organize material into a manageable number of categories. Match level of program with indicators appropriate for that level – remember that it is not possible for an evaluation to address all areas of interest.
- Come to agreement with stakeholders on the degree of incompleteness that is acceptable, given monetary and time constraints.
- Focus the scope of the evaluation on the crucial and practical.

In addition to talking with stakeholders, consider a variety of sources when clarifying the purpose of the evaluation, identifying issues, and developing the questions, indicators, and criteria.

- Examine various evaluation models and relevant available literature.
- Refer to professional standards and guidelines relating to the program area.
- Consult experts in the field.
- Use your professional judgment.

Indicators are observable phenomena that point toward the intended and/or actual condition of situations, programs, or outcomes. *An indicator is a marker that can be observed to show that something has changed or improved.* Indicators can help people notice changes at an early stage of a program's impact.

Criteria for Choosing an Indicator

- Is it measurable?
- Is it relevant and easy to use?
- Does it provide a representative picture?
- Is it easy to interpret?
- Does it show trends over time?

- Is it responsive to change?
- Does it have a reference against which to compare it?
- Can it be measured at a reasonable cost?
- Can it be updated?

Characteristics of Indicators

- Relevant to the objectives of the program to be evaluated
- Understandable, that is to say, simple and unambiguous.
- Realizable, given logistic, time, technical or other constraints
- Conceptually well-founded
- Limited in number and can be updated at regular intervals

Examples of indicators to evaluate the effectiveness of extension program focusing on agricultural production program include:

- Change in yield/unit
- Change in cost of production
- Change in input $/unit
- Change in $ per marketed unit
- Change in production quality and/or safety

Sometimes extension systems conduct evaluations of facilitation, human capacity building, and technical assistance. Indicators to evaluate such efforts may include:

- Number of staff receiving degree training or short-term training
- Number of technical assistance programs or services extended
- Quality of skill developed
- Improved job performance of staff members
- Improved morale of employees

STEP 5: IDENTIFY APPROACHES TO DATA COLLECTION

There are two basic types of data: quantitative and qualitative. Quantitative data tend to focus on numerical data and qualitative data are expressed in words.

Quantitative methods measure a finite number of pre-specified outcomes and are appropriate for judging effects, attributing cause, comparing or ranking, classifying and generalizing results.

Quantitative methods are:

- suitable for large-scale projects, especially those having a large number of participants;
- useful for judging cause and effect;
- accepted as credible; and
- applicable to or generalizable to a larger population.

Qualitative methods take many forms, including rich descriptions of people, places, conversations, and behavior. The open-ended nature of qualitative methods allows the person being interviewed to answer questions from his or her own perspective. Qualitative methods yield good evaluation data as long as they are unbiased and objective.

Qualitative methods are appropriate for:

- understanding the context in which a program takes place;
- Addressing complex problems and process issues;
- clarifying relationships between program objectives and implementation;
- identifying unintended consequences of a program;
- gathering descriptive information;
- understanding operations and effects of programs; and
- conducting in-depth analysis of program impacts.

As indicated earlier, program evaluators hold different epistemological perspectives and differ in their approaches to evaluation data collection. When selecting the appropriate methods, purpose of the evaluation, quality of evidence, and users of evaluation results should be considered as key elements.

The validity and reliability of the data collection instrument determine the quality of evidence, especially for quantitative methods. One of the most important steps in creating an evaluation instrument such as a survey is to establish its validity and reliability. According to Mueller (1986), validity and reliability are the most important criteria for assessing the quality of instruments.

Validity asks the question, “Does the instrument measure what it purports to measure?”

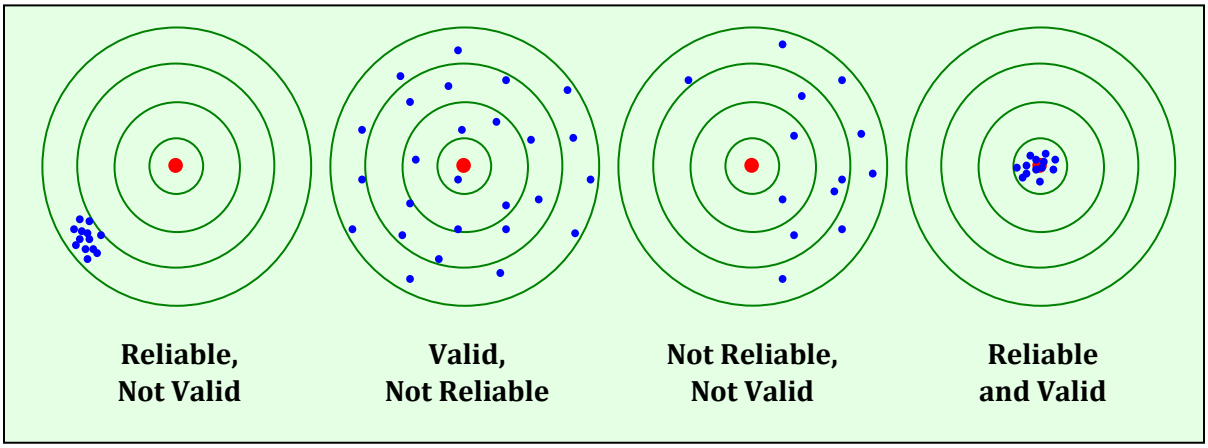
The following process is recommended to establish the validity of an instrument:

- Clearly define what you want it to measure (e.g., reactions, knowledge level, people involvement, behavior change).

- Prepare a draft of your instrument. Search for existing instruments related to your topic as a guide in developing your own instrument. You may use similar question formats and response categories.
- Recruit five to seven persons to serve as a panel of experts for reviewing your instrument in terms of content, format, and audience appropriateness. Remember that the members of the panel should be familiar with the purpose of the study. They should be able to judge whether you are asking the right questions to intended audience. Ask the panel to review the subject matter content of the instrument and give feedback.
- Revise the instrument by incorporating the suggestions offered by the panel.
- Field-test the instrument to test its suitability and clarity. Select about 10 persons who are similar to the target audience to participate in the field test. If possible, watch people complete the questionnaire. Watch for hesitation, erasures, or skipped questions. Seek verbal feedback after you have watched them complete the instrument. If some respondents appear confused or hesitant, ask why. Based on the feedback, revise your instrument.

Reliability asks the question, "Does the instrument consistently yield the same results with the same group of people under the same conditions?" Reliability looks for consistency, accuracy, and dependability of an instrument.

Figure 12: Illustration of Concepts of Validity and Reliability



Usually, reliability is established by conducting a pilot test. A test-retest method also can be used to establish reliability. This method involves administering the same instrument twice to the same group. Ask 15 to 20 persons having characteristics similar to the target audience to complete the entire instrument. After about two weeks, re-administer the same instrument to the same group of people. Compare responses on each question are compared in pairs (i.e., first-time and second-time answers from the same individual are compared). A high degree of agreement (70% or higher) between the paired scores across all questions and all respondents indicates that the instrument is reliable (Neito & Henderson, 1995).

In the diagram above, reliability and validity are assured if all dots lie inside the bull's-eye. This means that we are measuring the construct we want to measure and we are measuring it consistently.

***Internal consistency*:** Methods of assessing reliability based on internal consistency require that an evaluation instrument be administered only once. Multiple internal-consistency methods for establishing reliability are available. Some frequently used methods are:

- *Split-half procedure*: This procedure involves scoring two halves (usually odd items versus even items) of a test separately for each person, then calculating a correlation coefficient for the two sets of scores. The coefficient indicates the degree to which the two halves of the test provide the same results, and hence describes the internal consistency of the instrument (Fraenkel & Wallen, 1996).
- *Kuder-Richardson approach*: This approach utilizes three pieces of information about a test – the number of items in the test, the mean, and the standard deviation. It assumes that the items are of equal difficulty.
- *Alpha coefficient*: Cronbach's alpha is another procedure you can use to check for internal consistency of an instrument. This procedure is done by calculating reliability of items that are not scored right versus wrong. This procedure is appropriate to establish reliability of questions asked on a scale designed to measure reactions, attitudes, or perceptions (e.g., 1=strongly disagree, 2=disagree, 3=neutral/no opinion, 4=agree, 5=strongly agree).

A frequently asked question about reliability is, "What value of reliability coefficient is adequate to establish the instrument's reliability?" There is no hard and fast answer to this question. Consider what type and how important a decision has to be made using the results of the test. The more important the decision is to be made, the higher the reliability needed. Generally, an alpha value of .7 is considered as the cutoff for acceptable reliability. A low reliability may be acceptable if the measuring instrument has high validity (Kerlinger & Lee, 2000).

In summary, the evaluation plan for an agricultural extension program should be based on the outcome of a series of steps in the planning process. The evaluation should be planned at the time of developing the project itself, not as an afterthought. The standards and indicators should also be considered at the time of planning. The plan may include a series of formative studies for organizational learning, monitoring of program implementation, or impact assessment of major programs. Some of these evaluative studies may be conducted internally; others might involve external help. Program managers need to plan these studies at the onset of a project.

QUESTIONS FOR DISCUSSION

1. *Have you planned and implemented an evaluation for a program, project, or policy? Describe the nature of the evaluation you conducted and share it with your group.*
2. *What are the major challenges in designing a program evaluation for extension?*
3. *What is the role of stakeholders in program evaluation? How could their expertise and experience be integrated into program evaluation?*
4. *If you were to design an evaluation for an extension program, would you follow a qualitative, quantitative, or mixed method approach? Justify your answer.*
5. *Identify a program or project you would like to evaluate. Describe it briefly in three or four paragraphs. Include the following information:*

    What is the program or project?

    When did the project start?

    What are the goals/objectives of the program/project?

    What are the major activities of the program/project?

    Who are its audience (clientele or beneficiaries)?

    Who funded it, how much and for how long?

    Who is interested (or, who might be interested) in the evaluation of this program or project? What do they want to know about this project?

    Have you made any contact with the program/project staff about your interest in planning its evaluation?

    Have you conducted an "evaluability assessment" for this program or project?

This could be your individual evaluation project. Share your preliminary idea of your evaluation project with the group. During this training, you will refine your ideas in a more formal three to five page proposal.

CHAPTER V

Managing Evaluation Data Collection

STEP 6: SELECTING DATA COLLECTION TECHNIQUES

Evaluation data can be gathered from primary or secondary sources. Primary sources include original documents, the first reporting of facts, and the first grouping of the raw data. Secondary sources are materials that bring together facts from multiple primary sources.

Evaluation data generally are gathered from respondents (e.g., participants, facilitators, project managers, local residents, experts) by asking questions and collecting (orally or in writing) their deliberate responses. Sometimes, evaluation information is gathered from a subject by watching what happens in a specific context or situation (e.g., how participants react during workshops, how the nitrate level changes in a river system over time). Evaluation information also could be gathered from documents (e.g., training materials developed by project staff, minutes of meetings, correspondences, local newspaper articles).

Factors to consider in choosing the method:

- Resources available (staff support, time, money)
- Experience and expertise of evaluators
- Facilities at your disposal
- Sensitivity of the method to various kinds of errors

How do we decide which sources to use? It depends on several factors. First, we need to consider the information availability issue. Are the sources inexpensive to access? Do we have logistically easy access to them? Are human sources likely to cooperate? Second, we need to consider the credibility of the sources. Are they in the best position to report on the event? Are they likely to remember accurately? Is there reason to suspect that they might exaggerate or underplay the truth? Is their information likely to be incomplete? Do they possess the general background knowledge required to respond competently to the evaluation questions we want answered?

There is no one best method! We have to consider the relative merits of each method. Selection of the method should be influenced by the type of information desired, time availability, and cost. Many methods could be used, but you should choose those that provide the most useful information, those you and/or your staff have the skill to use, and those that are within your budget. Last, but not least, consider whether the information collected will be viewed as credible and accurate, and will be useful to your organization.

We must be precise about what we actually need to know. Don't be vague, biased, or non-critical. At the same time, we have to ensure that legal and ethical standards are maintained throughout the process of evaluation data collection, data analysis, and reporting.

Table 3: Common Evaluation Tools by Program Stage

Stage of program	Type of evaluative study	Typical questions answered	Example of evaluation tools/techniques
Project Planning Stage	Needs assessment Feasibility study Baseline study	What are the felt and unfelt needs of the audience? Can extension address these needs? Do they fit with extension's mission? Is the program/project socially, economically, environmentally feasible?	Focus Group Surveys and personal interviews Observation Content analysis Census and office records Economic analysis e.g., benefit/cost analysis, rate of return on investment)
Implementation Stage	Formative evaluation Program monitoring	Is the program meeting its objectives of intended outcomes? Is the audience satisfied with the program? Are the media delivering program messages?	Annual monitoring reports (e.g., staff time and activity reports, crop yield, seed cost) Adoption patterns for new technology Pre-post evaluative studies on knowledge, attitude and behavior change Customer satisfaction surveys Participant observation
Concluding or Results Stage	Summative evaluation	Has the program addressed the needs or gaps identified? Is the program achieving desired outcomes? Is the program cost effective?	Pre- and post-project data analysis Cohort studies Panel studies Surveys (e.g., personal interviews, telephone survey, mail survey, online survey) Economic analysis

SECONDARY DATA (EXISTING INFORMATION)

Census statistics about population growth, food production, economic and environmental conditions, or needs of specific groups of people are often available for public use and at minimal cost. This information is available for each county and data can also be subscribed to on a regular basis.

Information is obtained through document analysis (reviewing existing documentary information), such as letters, diaries, photographs, records, receipts, reports, proceedings of meetings or hearings, newspaper articles and editorials. These types of data provide insights into a program that cannot be observed in any other way.

> Before you start to collect primary data, determine what information already exists. Pre-existing information can be found in documents, reports, program records, historical accounts, minutes of meetings, letters, photographs, census data and surveys.

Existing information is useful for:

- establishing the need for a program: use census data, media feature stories, maps, or service and business statistics;
- describing how the program was carried out and who it reached: use program documents, log books, meeting minutes , enrollment records, and media releases; and
- assessing results: use public records, local employment statistics, agency data, and evaluation of similar programs.

Use of secondary data has many advantages. Many kinds of data are available and, in most cases, they are readily available. Data can be obtained with minimal cost and effort. Data can have high credibility.

Secondary data, however, have limitations. Generally data tend to be descriptive of the population and may not provide information to a particular community or geographic region of interest. Some figures may represent estimates or projections rather than actual counts. The data may not reveal values, reasons, or beliefs underlying the current trends. Further, information may present a biased view of reality.

DATA FROM TESTING AND EXAMINATION

Tests are useful tools for measuring an individual's knowledge, understanding, and ability to apply knowledge. They provide an indication of level of knowledge and other changes related to a particular program. They are relatively easy to implement and can be carried out in a group setting.

Basic Steps in Constructing a Valid and Reliable Knowledge Test

Step 1. Assemble a draft test consisting of questions that focus on the specific subject matter presented during the program.

Step 2. Request subject-matter specialists or experts to review the test items and suggest answers to ensure validity.

Step 3. Administer the test to a group of program participants.

Step 4. Review the responses or answers by looking for consistency and variability.

Step 5. Summarize and analyze the degree of comprehension achieved through the program.

Step 6. Perform a reliability test.

Note: Kruder-Richardson's KR-20 or KR-21 is used for dichotomous or binary scoring, such as for correct/incorrect responses and true/false responses. Cronbach's alpha is used to determine internal consistency of testing instruments that use Likert scales. Statistical software, such as SPSS, has scale functions that are easy to use and interpret.

Advantages of using tests:

- Tests can provide an indication of knowledge gain, attitude change and behavior modifications related to a particular program.
- Tests are relatively easy to implement.
- Tests can be carried out in a group setting.
- Tests tend to be low-cost to administer.

Limitations of using tests:

- Adults often resist attempts to test their knowledge.
- Valid and reliable tests require special skills and time to develop.
- Tests are not appropriate for less literate audiences.
- If the goal of the extension program is to bring about behavioral change, then tests are not a sufficient measure because knowledge gain does not necessarily cause or indicate a corresponding behavior change.

QUESTIONNAIRE SURVEYS AND INTERVIEWS

Surveys are a popular method of collecting evaluative data. Surveys are used to measure people's opinions, attitudes, beliefs, behaviors, reactions, and attributes in response to specific questions. They can provide distributions of some characteristics in a population. This usually can be accomplished by surveying only a portion of the people (or units) in that population.

Surveys have several advantages. They are moderate in cost and it is relatively easy to reach large numbers of the population. They allow for anonymity of responses. Evaluators can also ask fairly complex questions about respondents' attitude and behaviors. Data can be requested from records and other sources. Surveys can be conducted using mail, personal interview, telephone, Internet, or administered in a group setting, such as workshops or classrooms. Each method has its own advantages and disadvantages.

MAIL SURVEY

A mail survey is the most frequently used type of survey in evaluation of agricultural extension and rural advisory service programs and requires the fewest resources.

Advantages of using a mail survey:

- Can be used with a large sample size and/or with a widely dispersed population or one that is not accessible by telephone or personal interviewing.
- Provides a visual display of questions
- Is free of interviewer bias.
- Enables respondents to give thoughtful answers and control the pace of responses.
- Are relatively inexpensive.

Limitations of using a mail survey:

- The questionnaire must be short and carefully designed.
- The response rate depends on the number of contacts made with the respondent and the timing of the mailing.
- There is little control over the completeness of the responses.
- Those who reply may not be representative of the target population.
- Pretesting of the questionnaire is necessary to avoid costly mistakes.
- It requires a literate population and a reliable postal system.

Basic Steps in Implementing a Mail Survey

Step 1. Prepare survey material. Design a written questionnaire, using an identification number on each questionnaire to track returns. The appearance of a mailed questionnaire is of utmost importance. A mailed questionnaire must "sell" itself to the respondent in order to be returned. Therefore, considerable care should be taken in designing the format of the questionnaire.

- A simple booklet can be constructed by folding an 8 ½ by 11-inch paper in half.
- Make questions fit the page so that the respondent does not need to turn the page to answer a question.
- Provide easy-to-follow directions on how to answer the questions.
- Arrange questions and answers in a vertical flow. Put answer choices under, rather than beside, the questions.

Step 2. Pretest instrument to assure validity and reliability.

Step 3. Select survey population, develop sampling frame, and determine sample.

Step 4. Develop a mailing schedule: a) Two weeks before mailing the survey, send an advance letter; b) mail the questionnaire, including a cover letter and a stamped, self-addressed envelope; c) send a postcard a week or so later, thanking those who responded and reminding those who did not to return their surveys; d) three weeks after mailing the first questionnaire, send a follow-up letter stating that a response has not been received; include a replacement questionnaire and a stamped, self-addressed envelope.

PERSONAL INTERVIEW SURVEY

Personal or face-to-face surveys are conducted by talking individually to respondents and systematically recording their answers to each question. Personal interview surveys administered in person are common method of survey data collection in the developing countries.

Advantages of a personal interview survey:

- It can be used with a highly dispersed population.
- It is suited for populations for which a representative sample cannot be drawn.
- It can be used where there is a low literacy rate.
- There is a high degree of control over who answers the survey.
- The interviewer can increase the willingness of respondents to answer questions.
- Visual aids can be used to facilitate understanding of survey questions.
- Questions can be fairly complex.

Limitations of a personal interview survey:

- It can be expensive and time-consuming.
- Interviewers must be carefully selected and receive adequate training.
- It requires a good supervisor.
- It requires appropriate training and close supervision.

Basic Steps in Implementing a Personal Interview Survey

Step 1. Develop survey material, including:
- an advance letter if names and addresses are available;
- an introductory letter explaining the purpose of the survey;
- an interviewer's instruction manual;
- sampling information for interviewers; and
- the questionnaire.

Step 2. Identify and train a staff of interviewers.

Step 3. Mail letters/circulars describing the survey and telling them to expect a visit from an interviewer. Also, Notify public officials about the survey.

Step 4. Conduct interviews. A supervisor should be available while the survey is being carried out to handle any problems that may arise.

Step 5. The supervisor should meet regularly with interviewers to answer any questions interviewers may have. Costly errors, misunderstandings, and cheating by interviewers can be detected at this time.

Step 6. After interviews are completed, the questionnaires are returned to the survey supervisor.

Interviewer bias is a major challenge in collecting quality data. Although it may be impossible to eliminate interviewer bias, it can be minimized.

- Training of interviewers prior to data collection.
- Being familiar with the research instrument.
- Maintaining a neat and professional appearance.
- Following sampling instructions properly.
- Being honest with the respondent.
- Asking questions exactly as written.
- Recording responses accurately.
- Checking for completeness of survey information.

Initiating contact:

- Introduce yourself; show your identification/ID or credentials.
- Remind respondent of the notification letter he or she received a few days earlier.
- Explain the purpose of the survey.
- Assure respondent that his/her answers are voluntary and will remain confidential.
- Explain how respondents were chosen.
- Explain "burden" and benefits (e.g., how long survey will take, how results will be used, incentives, potential benefits to them or the community)

Guidelines for interviewing:

- Select a mutually convenient time for interview.
- To avoid distractions, try to conduct the interview without distractions.
- Establish rapport by expressing appreciation of the respondent's responses and willingness to participate.
- Read questions as they appear in the questionnaire and record answers accurately.
- Do not express your opinions.
- If an answer to an open-ended question is incomplete or appears irrelevant, probe to get a clearer response.
- If a respondent refuses to answer a question, do not insist on getting an answer. It may jeopardize the entire interview, and doing so is inconsistent with the voluntary nature of their participation in the study.

TELEPHONE SURVEY

A telephone survey consists of a written questionnaire that is read to selected individuals over the telephone. The survey sample is often selected from a telephone directory or other lists. People on the list are interviewed one at a time over the phone. Responses are recorded by the interviewer either on paper or a digital tool (e.g., computer).

Basic Steps in Implementing a Telephone Survey

Step 1. Find suitable facilities and equipment necessary to implement the survey. Computer-aided telephone survey software is available. Make sure data collection staff is familiar with the survey software.

Step 2. Decide on a sampling design, including the method of respondent selection within a sampling unit. Choose the method to generate a pool of telephone numbers that will be used in sampling.

Step 3. Prepare survey material: An advance letter if names and addresses are available; the questionnaire (keep it short by asking only necessary questions); a cover sheet to record identification number and a call-sheet; and help sheets for the interviewer.

Step 4. Train interviewers on: background information about the survey; basics of telephone interviewing; how to use equipment; and how to fill out questionnaires and call-sheets.

Step 5. Develop an interview schedule. Assess when you will be likely to contact respondents, during working or non-working hours. (This will vary by country, occupation, local customs, etc.) For most surveys, approximately 30 minutes is sufficient to complete an interview. Decide how to handle refusals.

Step 6. Make calls. Decide on the number of calls to make to each number. In some surveys, six to seven calls are customary. Make callbacks.

Note: *Telephone questionnaires depend on oral communication, so special attention must be paid to designing a questionnaire that will assist the interviewer as much as possible in holding the respondent's attention. Design and construction of the questionnaire are based on utility rather than aesthetics.*

INTERNET OR ONLINE SURVEY

Internet or online surveys are similar to personal surveys except the mode of communication is via computer using the Internet. Internet or online surveys were a novelty until the late 1990s. Only a few households had e-mail connections and they were very slow with dial-up modems. Today, many households have high-speed internet access. University students, school teachers, professional workers and employees—all have access to the Internet and this method is gaining popularity.

Compared to mail surveys, Internet and online surveys, e.g., SurveyMonkey, have more power and flexibility due to the potential for incorporating built-in features such as:

- dropdown menus;
- slider scales;
- pictures or photos to illustrate scales;
- color, animation, sounds;
- video; and
- feedback screens and hotlinks.

GROUP-ADMINISTERED QUESTIONNAIRE

A group-administered questionnaire is handed directly to each participant in a group at the end of a workshop, seminar or program. Respondents answer the questions individually and return them to the person conducting the evaluation.

Basic Steps in Implementing a Group-administered Survey

Step 1. The questionnaire is prepared following the guidelines for constructing a survey instrument. However, the objectives and instructions for completing the questionnaire are explained to the participants by the instructor, supervisor, or agent. He or she also should tell attendees that their participation is voluntary, and assure them that their responses will remain confidential.

Step 2. The questionnaire is distributed to each participant to be filled out individually.

Step 3. The questionnaire is collected and checked for completeness.

CREATING QUALITY SURVEYS BY AVOIDING ERRORS

Sometimes surveys produce inaccurate results due to data collection errors. Accuracy means that survey results closely represent the population from which the sample has been drawn. Inaccuracy can be caused by several types of errors, including coverage error, sampling error, selection error, frame error, non-response error, or measurement error.

Although not all error can be eliminated, the evaluator can minimize the potential for various types of error by taking specific actions throughout the evaluation development and implementation process. Table 4 summarizes the type of errors and how to minimize them.

Table 4: Types of Survey Errors and Ways to Control Them

Type of Error	Cause of Error	Control of Error
Coverage error	The sampling frame does not include all units of the population.	Redraw list from which the sample is drawn to include all elements of the population.
Sampling error	A subset or sample of all people in the population is studied rather than conducting a census.	Increase the size of the sample; use random sampling; purge list of duplicate entries.
Selection error	Some sampling units have a greater chance of being chosen than others are.	Use random sampling
Frame error	List is inaccurate or some sampling units are omitted.	Use up-to-date, accurate list.
Non-response error	Subjects can't be located or they fail to respond. Sometimes people who do respond to a survey are different from sampled individuals who do not respond.	Dillman (1994) suggests the use of a social exchange concept to improve response rate, i.e., increase respondents' perception of possible rewards (e.g., provide token incentives), decrease perceived costs (e.g., time) and encourage trust (e.g., promote trust by showing trustworthy sponsorship) so that rewards outweigh costs. Miller and Smith (1984) suggest the following strategies: • Compare early and late respondents. If no difference is apparent, results can be generalized. • Contact 10% of non-respondents and compare these data with the respondents. If no difference is apparent, results can be generalized.

Type of Error	Cause of Error	Control of Error
		• Compare respondents to non-respondents on known characteristics. If no difference is apparent, the results can be generalized.
Measurement error	A respondent's answer is inaccurate or vague. This may be due to: unclear questions or instructions; socially correct responses, respondent not knowing the correct information, or deliberately lying.	• Choose appropriate method of data collection for your evaluation. • Write clear, unambiguous questions that people can and want to answer. • Train your interviewers carefully. • Use valid and reliable instruments.

GENERAL GUIDELINES FOR QUESTIONNAIRE DESIGN:

The overall aim of questionnaire design is to solicit quality participation. Response quality depends on the trust the respondent feels for the survey, the topic, the interviewer and the manner in which the questions are worded and arranged. Consider whether the questionnaire is going to be mailed, given directly to respondents, used in a telephone survey, or used in personal interviews. Before you begin, it is essential to know what kind of evidence you need for the evaluation, how the data will be analyzed, and how the information will be used.

Before you begin…

- Make a list of what you want to know and how the information will be used.
- Check to make sure the information is not already available somewhere else.
- Eliminate all but essential questions.
- As you write questions, try to view them through the eyes of the respondents.

Guide to Writing Questions and Designing a Questionnaire

1. The title and accompanying graphic of the questionnaire should appeal to the respondents.
2. The type used should be large and easy to read.
3. The questionnaire should appear professional and easy to answer.
4. The introduction should identify the audience, describe the purpose of the survey, and give directions about how to complete the questionnaire.
5. Questions should not appear crowded. Each question should be numbered and sub-parts of a question should be lettered.
6. Questions should be arranged in a logical order, with general questions preceding more specific ones. Easy-to-answer questions come first, followed by increasingly complex, thought-provoking, or sensitive questions. Personal or potentially threatening questions should be placed at the end. A request for demographic information should be included near the end of the questionnaire.
7. Sufficient space should be left for answering open-ended questions.
8. Clearly indicate where branching occurs and where general questions resume.
9. Key words should be boldfaced or capitalized to minimize the possibility that they are misread.
10. The questionnaire should end with a "Thank You."

WRITING QUESTION ITEMS

The questions used in a questionnaire are the basic components that influence the effectiveness of your survey. Writing good questions is not easy and usually takes more than one try. Consider what information to include, how to structure the questions, and whether people can answer the questions accurately. Good survey questions are focused, clear, and to the point. Questionnaire writers should consider question specificity, meaning, length, and potential bias when writing individual questions.

Specificity: Every question should focus on a single, specific issue or topic.

> **Poor:** Which brand of coffee do you like best?
>
> **Better**: Which one of these brands are you most likely to buy – Folgers, Starbucks, or Maxwell House?

The objective of the questions above is to measure consumer purchasing preference. The first question lacks focus; consumers may like a particular brand, but may not buy it because of its high price.

Meaning: The meaning of the question must be completely clear to all respondents. Clarity ensures that everyone interprets the question the same way.

> **Poor:** When was the last time you went to the doctor for a physical examination on your own or because you had to?
>
> **Better:** How many months ago was your last physical examination?

The first question could be interpreted in weeks, months, years, or by date.

Length: Keep questions as short as possible. Short questions are easier to understand and answer, and are less subject to error by interviewers and respondents. Long questions are more likely to lack focus and clarity.

> **Poor:** Can you tell me how many children you have, whether they're boys or girls, and how old they are?
>
> **Better**: What is the age and sex of each of your children?

A respondent may answer the first question ambiguously. For example, "I have two boys and a girl. They are 5, 7, and 10 years old." It is not possible to determine the ages of each child from this response.

Bias: Questions should be written to avoid bias.

> **Poor:** Is it true that our agents always work long hours?
>
> **Better:** On average, how many hours per week do extension agents work in their jobs?

TYPES OF INFORMATION

Questions can be formulated to elicit four types of information: 1) knowledge, 2) beliefs, attitudes and opinions, 3) behavior, and 4) attributes. Any one, or a combination, of these types can be included in a questionnaire.

Knowledge questions include what people know and how well they understand something.

Example: What is the major cause of accidental deaths among children inside the home?

Belief, attitude and opinion questions solicit people's perceptions, their thoughts, their feelings, their judgments, or their ways of thinking about the topics or issues of interest.

Example: Should the Clearwater Regional Education Center in Minor County continue to offer college-level and/or continuing education courses and programs? Why/why not?

Behavioral questions ask people about what they have done in the past, what they do now, or what they plan to do in the future.

Example: Have you or your family ever taken classes at the Clearwater Regional Education Center in Minor County?

Attributes are a person's personal characteristics, such as age, education, occupation, and income. Attribute questions ask respondents who they are, not what they do.

Examples:

Where do you currently live?

How many children do you have?

What percentage of your household income comes from off-farm employment?

TYPES OF QUESTIONS

There are basically two distinct types of questions asked in a survey – closed-ended questions and open-ended questions.

Closed-ended questions

Closed-ended questions have pre-determined categories of responses from which the respondent can choose. When asking closed-ended questions, make sure to include all alternative response categories. Sometimes an "other" category is provided, with space for respondent to specify. For some questions, respondents must choose only one response; for others, they may select as many as are relevant.

Examples of Closed-ended Questions

1. Have you or members of your family taken classes at the Extension Center this year?
   __Yes __ No

2. To what extent do you agree or disagree with the new land tax policy? Circle one.

   Strongly disagree Disagree Neither Agree Strongly agree

3. Approximately, how much did you spend on fertilizer during 2011?

   __None __$1 -100 __$101-200 ___$201 and more

Open-ended Questions

Open-ended questions allow respondents to answer in their own words rather than select from predetermined answers.

Examples of Open-ended Questions

1. How do you plan to use the information acquired during this training?
2. What do you think should be done to improve the 4-H program in this county?
3. How much did you spend on fertilizer in 2011? _____

PRE-TESTING EVALUATION INSTRUMENTS

Pre-testing means trying the method and the instruments before actual data collection in the field. It is usually associated with quantitative methods, though qualitative and participatory methods can be pre-tested as well. The pre-testing process avoids costly errors and wasted effort. When possible, pre-testing should be done in circumstances similar to those anticipated during the evaluation. If feasible, use the same sampling plan you will use during the evaluation to select a mini-sample.

In pre-testing, we ask questions such as**:**

- Are the issues to be discussed, the questions to be asked, and/or the words to be used clear and unambiguous?
- Is the technique or instrument appropriate for the people being interviewed or observed?
- Are instructions for the interviewer or observer easy to follow?
- Are the techniques and/or forms for recording information clear and easy to use?
- Are procedures standardized?
- Will the technique or instrument provide the necessary information?
- Does the technique or instrument provide reliable and valid information using the criteria of the chosen data collection approach?

You may find that you have to modify the technique or instrument after field testing. If extensive revisions are made, a second field test may be necessary.

Students and beginning practitioners frequently ask, "which one is better, mail, telephone or online/Internet?" Advantages, disadvantages, and basic steps for implementing each are described below.

Which one to choose… mail, personal interview, telephone or Internet survey?

Coverage is an issue for online and telephone surveys. Not all people have access to the Internet or telephone service. Mail surveys may not be appropriate for contexts in which mailing addresses or postal services are not reliable, or target segments of the population cannot read or write. The context of your evaluation, the nature of evaluation questions to be answered, and the resources available for data collection may dictate the choice of method. Each method has advantages and disadvantages. These must be considered, along with characteristics of the context, costs, time available, purpose, and other factors, when choosing an appropriate survey type.

Mail is the method of choice when: (a) size of sample is large, (b) visual display of questions is needed, (c) educational level of respondents is high, (d) respondents are dispersed in a large geographical area, and (e) the budget is low. If designed properly, the surveys can generate valid and reliable information. A mail survey, however, should be avoided if the target population has low education, survey questions are open-ended, postal services are inadequate or weak, or sampling frames are inadequate or not available.

Online or Internet survey is better when: (a) e-mail addresses of respondents are available, (b) respondents have access to the Internet, (c) sample size is large and the budget for data collection is low, and (d) survey needs to be completed in a short time.

Telephone survey is the method of choice when: (a) respondents are widely dispersed geographically, (b) speed in data collection is essential, (c) sample size is small, and (d) cost is not a big factor. Telephone surveys may yield a higher response rate than mail surveys. To some extent, interviewers can explain questions not understood by the respondents. Telephone surveys, however, should be avoided if we need to ask long and complex questions and/or bias against people without telephones cannot be tolerated. The cost may be higher than a mailed questionnaire, it requires good interviewing skills, and there is a natural bias in favor of those with listed numbers and those who are usually in their homes. It requires clear and simple questions. If a respondent is unfamiliar with the organization or caller, there might be indifference and/or poor cooperation.

Group administered surveys are used when: data are to be gathered under group situations like at the end of workshop, seminar, classroom, etc. This approach has two major advantages: (a) there is little or no cost in reaching respondents and (b) the purpose behind asking for the information can be clearly explained. The disadvantages include: (a) limited generalizability of information to a larger population, (b) it takes time away from the regular program, (c) group mood or setting at the time may affect responses, and (d) it does not allow for long term reactions and changes.

Evaluators could use "mixed-mode surveys" to collect some data by mail and some by telephone when (a) one method won't get an adequate response rate and/or (b) faced with sampling problems. Dillman (1994) warns that mixed-mode surveys should be avoided when key evaluation questions involve attitude and/or social desirability.

As program evaluators, we should put special effort into holding all of the above four types of errors to acceptable levels while designing the evaluation.

FOCUS GROUP

Focus groups often are used in marketing research to find out what a particular component of the public needs or wants, and what they will consume. In recent years, this technique frequently has been used to identify community needs and issues, to obtain citizens' perceptions on a defined area of interest, to generate program alternatives, and to assess the impacts of a particular program on individuals and communities. Focus group interviewing uncovers information on human perceptions, feelings, opinions, and thoughts.

A focus group typically is composed of seven to ten participants, with members selected because they have certain characteristics in common, or interests that relate to the topic of the focus group (Krueger and Casey, 2000).

A focus group is a small group, typically consisting of 7 to10 people who are relatively homogeneous, which is selected to discuss a specific topic in a non-threatening atmosphere. The focus group is moderated and recorded by a skilled interviewer. A focus group identifies community needs and issues, citizens' attitudes, perceptions, opinions on specific topics, and impacts of a particular program on individuals and communities.

Focus groups should be conducted by a skilled interviewer. The interviewer should create an open environment in the focus group that nurtures different perceptions and points of view, without pressuring participants to vote, plan, or reach consensus. Krueger and Casey (2000) suggest that the discussion needs to be relaxed, comfortable, and potentially enjoyable for participants as they share their ideas and perceptions. The group discussions should be conducted several times with similar types of participants to identify trends and patterns in perceptions. Careful and systematic analysis of the discussions provides clues and insights as to how a product, program, or service is perceived.

HOW TO BEGIN A FOCUS GROUP DISCUSSION

The first few moments in a focus group discussion are critical. In a brief time, the moderator must create a thoughtful, open atmosphere, provide the ground rules, and set the tone of the discussion. Much of the success of group interviewing can be attributed to the development of this open environment. The recommended pattern for introducing the group discussion includes: the welcome, the overview and topic, the "burden for participation," the ground rules, reaffirmation of confidentiality and voluntary nature of participation, and an invitation for participants to introduce themselves.

A program evaluator may consider the following guidelines (Adapted from Krueger & Casey (2000)):

- Consider your purpose: Why do you want to conduct a focus group interviews? Who are the users of this information? Why do they want the information?
- Develop a tentative plan, including resources needed.
- Identify the questions to be asked in the interview, including both primary and follow-up guiding/probing questions.
- Arrange a suitable meeting place in a convenient, preferably neutral, location. It could be a meeting room in the Courthouse, at a local restaurant, or school.
- Identify the audience who will be interviewed. Invite them well in advance. Explain to them the purpose of the meeting and how they can contribute. Reconfirm their availability to participate in the session.
- Identify a trained moderator (and an assistant) to conduct the focus group interview. The moderator must be mentally alert and free from distractions. He or she should help create a warm and friendly environment.
- Arrange the meeting room for the interview. Check the seating arrangements.
- Conduct focus group interviews. The moderator should: again explain the purpose and get written confirmation of consent; reassure participants about the voluntary nature of the study and the confidentiality of their responses; tape record the session; and guide discussion.
- Immediately after the interview, the moderator and assistant moderator discuss common experiences and perceptions that surfaced during the interview. They should review the tape together before the next focus group is conducted.
- Transcribe the taped discussion, then summarize what was said by the participants. Identify and analyze emergent themes. Analysis and interpretations should focus on meaning: What do the findings mean to you? Are the findings of value to the stakeholders? What recommendations are in order?
- Prepare a short report and share the findings with your stakeholders.

HOW TO WRITE AND USE QUESTIONS IN A FOCUS GROUP

Carefully prepare primary and probing guiding questions.

Identify potential questions. Five types of questions are:

- Opening questions (round-robin)
- Introductory questions
- Key questions
- Transition questions
- Ending questions

Use open-ended questions to stimulate discussion.

What did you think of the program?

Where do you get new information about _____?

What do you like best about the proposed program?

Avoid dichotomous questions (those that can be answered with “yes” or “no”).

“Why” questions are rarely asked.

“Why” questions can make people defensive and feel the need to provide an answer.

When you ask “why,” people usually respond with attributes or influences.

It’s better to ask, “What prompted you?” or “What features did you like?”

Use “think back” questions that remind respondents of an experience rather than asking them to speculate on the future.

During discussion, ask uncued questions first, cued questions second.

Cues are the hints or prompts that help participants recall specific features or details. These are often called “probing questions.”

Focus the discussion by using a sequence of questions that proceeds from general questions to those focusing on specific topics or issues of concern.

RAPID RURAL APPRAISAL

Rapid Rural Appraisal (RRA) is an assessment approach that involves multiple data collection techniques that are quick, flexible, and adaptive, yet relevant. The approach aims to incorporate the knowledge and opinions of rural people in the planning and management of development projects and programs. Usually, a multidisciplinary team of experts visits a community to learn about local people's situations, experiences, and problems from a local perspective. The team may use key informant interviews, observations or check lists, focus group interviews, a nominal group technique, and/or other group methods to solicit ideas, opinions and perspectives of the local people.

RRA can capture more accurate information than surveys. For example, RRA can be used to gather comprehensive information about farming conditions such as the following:

- Crops grown (by season)
- Land use intensity (e.g., cropping system, inter-cropping)
- Farming system (e.g., crop-livestock, use of agro-forestry)
- Soil types and soil conditions (e.g., degree of salinity, water-logging, drought)
- Land ownership pattern
- Number of plots owned, distances from farmstead to home or between plots
- Crop yields
- Agricultural practices:
  - Land preparation
  - Fertilizer application
  - Weeding
  - Irrigation, drainage
  - Transportation and marketing

Basic Steps for Implementing a Rapid Rural Appraisal

Step 1. Identify goals of RRA and develop questions to ask.

Step 2. Form a multidisciplinary team and a visit schedule.

Step 3. Identify possible sources of information.

Step 4. Review existing documentation.

Step 5. Identify, adapt, and/or create data collection and recording methods.

Step 6. Adjust questions, sources of information, and approaches, as needed.

Step 7. Plan when and where to visit, and whom to contact.

Step 8. Begin data collection while remaining flexible to the local situation.

Step 9. Record data, as collected, in a systematic fashion.

Step 10. Continually analyze data by verifying responses, deepening understanding, and making distinctions and connections between responses.

CASE STUDY

A case study provides in-depth information on a single unit, project, or organization. Using a systematic process, the evaluator captures the total essence of a situation through personal discussion, interaction, observation, and/or review of existing documents. Yin (1984) describes case studies as *explanatory, descriptive, or exploratory*.

- Exploratory case studies focus on information as a prelude to a more in-depth evaluation. This *can help identify performance measures or pose hypotheses for further evaluation.*
- Explanatory case studies *can measure causal relationships.* They seek to explain "how" and "why" something happens and what could possibly make them happen.
- Descriptive case studies are used to describe the *context in which a program takes place and the program itself.*

Case studies may also take the form of:

- Comparative case studies which are used to compare program processes and impacts of two or more cases or programs.
- Pre- and post-case studies which examine and describe the situation before and after a program or event takes place.
- Longitudinal case studies which look at a case at multiple times over the course of a program or project.

SEMI-STRUCTURED INTERVIEWS WITH KEY INFORMANTS

This method is based on obtaining information, over time, from program participants or residents who are in a position to know the program or the community well. Key informants in a community may include a school principal, local leaders, Church officials, local business leaders, and members of service clubs such as Lions, Kiwanis, or Rotary Clubs. These people could provide fairly representative information on how a program or project is serving its intended beneficiaries. The evaluator should, however, recognize the limitations of this approach in that it does not use random selection of subjects and, thus, is subject to information bias and lack of representativeness.

Semi-structured interview is a method of asking open-ended questions with key informants on a specific topic. Probing techniques are used to solicit in-depth answers and raise new topics that reflect the people's perspectives, beliefs, attitudes and concerns.

GUIDELINES FOR SEMI-STRUCTURED INTERVIEWING

- Identify topics and develop open-ended questions.
- Select respondents following the chosen sampling criteria.
- Remember that your appearance and mannerisms have an impact on the interview. Dress appropriately and speak in a non-threatening manner and use easy-to-understand terms.
- Conduct the interview or select a mutually convenient time to return.
- To avoid distractions, try to conduct the interview without an audience.
- Explain the purpose of the interview to the respondent and remind participants that the interview is voluntary and his/her responses are voluntary and will be kept confidential.
- Establish rapport by beginning with a general conversation on a neutral subject that might interest the respondent and share some personal background and express appreciation for the respondent's responses.
- Begin with simple questions that do not require long answers or a lot of reflection, then move on to more complex and sensitive questions.
- Record answers verbatim.
- Do not express your personal opinions.
- If an answer is incomplete or appears irrelevant, probe to get a clearer response.
- If a respondent refuses to answer a question, attempt to get an answer but avoid doing something that might jeopardize the interview.

PARTICIPANT OBSERVATION

Developed by anthropologists, participation observation is a method that is well described by its name. Rather than remaining detached, the participant observer lives with, eats with, works with, plays with, and may even join in rituals with the people he or she is studying. Participant observation entails gathering information about behavioral actions and reactions through direct observation, interviews with key informants, and participation in the activities being evaluated.

True participant observation requires the investigator to immerse him/herself in the life of the community being studied. This method is especially useful in the assessment of long-term effects on local residents of a new policy or development program. It is useful in determining reasons for community conflicts or misunderstandings, assessing community needs and problems, and finding acceptable ways of involving people in problem solving.

Participant observation alone will rarely provide enough information for a program evaluator. More detailed information usually must be elicited by interviewing informants. Such interviews may be particularly valuable to learn about local peoples' beliefs, values, motivations, power relationships, etc. The observer requires strong observational skills to document the complex human behavior because the quality of information is subject to biases of the observer.

There are some significant ethical issues involved with the study of other humans through participant observation. Many people do not enjoy being observed; it may make them feel self-conscious, awkward, or embarrassed. People have a right not to be observed if they do not want to be. Therefore, it is important to follow some general guidelines while practicing participant observation.

Whenever possible, ask people's permission to observe them. You can say something as simple as, "I'm very interested in learning about such-and-such from you. May I write down a few notes about our conversation?" You should assure them that their names will never be used in the report.

In some cases, it will not be possible to ask permission. There may be too many people, the action may be temporary, or the people may be at some distance. In these cases, observe and record only behavior that is enacted in public.

If anyone objects to your observation or data recording, you must respect their wishes and stop your activities immediately.

GENERAL GUIDELINES FOR ENGAGING IN PARTICIPANT OBSERVATION

Participant observation (PO), as used in evaluation, is motivated by the need to solve practical problems, not to construct theory. Therefore, the evaluator using this technique should enter the field with an initial conceptual framework (Casely, David, & Kumar, 1988.) The framework should include preliminary issues and the possible relationships among them.

- Define the main concepts in the framework, e.g. learning style, leadership.
- Identify sources of information.
- Select the site in which participant observation is to be carried out. Selecting two or more sites allows for comparative analysis of data. An informal sampling technique is used in PO. The site selected should be representative of the type of program or organization being observed, the organization must be willing to accept the PO evaluator, and the PO evaluator must be able to enter into activities under observation. Timing is crucial for one-time activities, seasonal events, or those having a daily routine.
- Arrange for access and develop arrangements for maintaining confidentiality.
- Assemble tools for observation: checklist, pen, camera, tape recorder, etc. How data is recorded depends on the situation. You may want to take notes on the spot or you may want to make notes after completing your observations. Photographs and recording devices assist in recording, but in some instances may be intrusive and influence the situation being observed.
- Begin observation. You do not need to observe everything that is going on, but rather should focus only on those aspects of the activity pertinent to the evaluation.

ECONOMIC ANALYSES FOR EVALUATION

Economic assessment is the application of economic principles and models to the evaluation of agricultural extension and development activities. Focusing on costs and benefits, economists distinguish between two broad types of evaluation: **ex ante** (before the project begins, to help select among alternative agricultural extension activities and allocate resources to them) and **ex post** (after an agricultural extension project is completed, to assess its results). Economic evaluation can assist in planning extension programs, in estimating extension pay-offs (past or future) and in guiding extension policies. Three methods are commonly used (adapted from Falconi, 1993):

Scoring methods involve the identification and weighting of several, mainly economic, criteria to allow commodities or research areas to be ranked on the basis of a composite score. They lend themselves to group work and active participation of researchers and managers in priority-setting. Economic criteria such as value of production, expected yield changes, and economic efficiency are considered.

The *economic surplus approach* is widely used to calculate benefit-cost ratios, internal rates of return, and net present values of benefits generated by agricultural extension.

Econometric methods can be used to estimate production, supply, cost, or profit functions providing information for decision making. Econometric methods are more accurate than other methods at assessing the contribution of extension programs to changes in total output, provided that reliable historical data are available and the analyst is a knowledgeable econometrician.

Generally, social scientists having economics backgrounds are engaged in impact evaluation. Valuing the economic impacts of extension programs is not simple. Often, economists depend on secondary data, such as census records, to track impacts. They also utilize surveys to gather information for determining impacts. Richardson and Moore (2000) offer various ways of valuing impacts. These include the following:

Reduced cost: This method measures the money saved by a participant. For example, a farmer adopts a less expensive pest management practice after attending an Integrated Pest Management (IPM) training program. The reduced cost is equal to the cost of regular pest management practice minus the new IPM practice.

Increased income: This method compares the income of a program participant before and after the program.

Savings: This method computes the amount of savings or increased savings attributable to an extension program as experienced by participants before and after the program.

Increased productivity: This method computes economic value by measuring the increase in productivity by the same number of workers or units of production due to adoption of a new practice as a result of participating in an extension program. For example, extension programs teach farmers how to use a new technology. The higher profits from using the new technology minus the cost of buying the technology equals increased productivity.

Value added: This term means that a product is used in a new way that is more profitable. For example, a program that teaches fruit farmers to make jelly is adding value to the fruit. The profit from selling jelly minus the profit made from selling the fruit equals the **value added**.

Expected values: This method estimates how much income a new business will have. Banks rely on this method when deciding to make a loan. The income of similar businesses is used to estimate the income of the new business. For example, extension agents can use this method to estimate the

value of an extension program that teaches participants how to start a small business. The **expected values** of the businesses started could be the value of the program.

Alternative opportunity cost of capital: Extension programs can teach participants how to make more money from existing capital. For example, land could be used to grow a higher value crop. The income from the higher value crop is compared to the income from the lower value crop to estimate the economic worth of the program.

Willingness to pay: The willingness of clients or consumers to pay for some item or service may be considered an economic benefit when this willingness exceeds what would be considered a standard norm for a product or service.

Multiplier effect: This method is commonly used in economic development. It estimates the multiple effects of increasing income. For example, increased income is spent with local merchants who then buy more goods. The effect of the money is greater than simply increased income for the program participant.

Non-market benefits: These benefits do not have monetary value, but they are important for quality of life. Examples include leadership skills, educational exposure, improved attitude, or self-esteem. It is important to remember that extension programs are valuable for non-monetary reasons.

Benefit/Cost Analysis

Benefit/cost analysis typically is viewed as an alternative to program evaluation. However, it can also be seen as an extension of the evaluation process. As such, benefit/cost analysis provides a means to systematically quantify and compare program inputs to program outcomes in monetary terms. Valuing both benefits and costs in monetary terms allows them to be compared directly to determine the net impact of the program, make comparisons between alternative programs or projects, assist in program planning, advance organizational accountability, and /or expedite program support.

Steps to Benefit/Cost Analysis

Step 1: Develop a list of costs and benefits from various sources. Program costs include direct costs, implied or indirect costs, and implicit or assumed costs. Include program descriptions, professional literature, your own knowledge, and information compiled during initial phases of analysis. Program benefits are the positive outcomes or consequences resulting from the program or project. They include direct benefits and those that accrue over time. When determining costs and benefits, make sure that costs and benefits are measured at the same level.

Advantages of benefit/cost analysis:

- It has high credibility as a source of information.
- It is useful in justifying budgets, demonstrating the value of a program, and/or assisting in getting the most outcomes possible from program inputs.
- It yields useful information for donors and funders.

Limitations of benefit/cost analysis:

- It may be difficult to quantify costs or benefits in monetary terms.
- It may be difficult to account for opportunity costs, hidden costs, and/or assumed costs.
- It may be difficult to account for indirect benefits of the program.
- Bias may occur when assigning monetary value to costs and benefits.
- Bias may occur through underlying and untested assumptions.

A. **The cost equation: Cost = L + K + I – i**

L = labor: The cost per hour for labor, including salary and fringe benefits. Fringe benefits vary but normally fall within 22 to 35 percent of full salary. The complete labor hourly formula (L) is: (S+S.35)/260/8 where S = salary and S.35 = 35% fringes, 260 = workdays per year, and 8 = hours per workday.

K = direct costs: Direct program costs budgeted for, or assigned to, the program (e.g., supplies, correspondence, communications, travel and per diem expenses, equipment, and audiovisuals). If costs are shared between projects, the total is calculated from a cost/share equation. Opportunity costs are defined as opportunities that participants have lost to participate in the program. Opportunity costs are included in direct costs to the participants, the presenters or the stakeholders, depending on the level of analysis.

I = indirect costs: Costs indirectly associated with the participants but directly associated with the program (e.g., administrative costs such as facility rental, photocopying, report costs, telephone, and prorated equipment and supplies costs).

i = discount amortization: Measurable returns over time (both positive and negative). Discount amortization is not included if returns cannot be traced over time.

B. **The benefit equation: B = Cr + DB + IB**

Cr = cost reductions attributable to program activities.

DB = direct benefits: the primary outcomes experienced by participants and others directly involved in the program. They are typically derived from program objectives.

IB = indirect benefits: secondary or intangible outcomes of the program or project experienced by participants, non-participants or society in general. These outcomes or consequences can be positive or negative.

Step 2: Compare costs with benefits, either directly by subtracting costs from benefits or as a ratio of benefit to cost. The first equation provides a means of comparing costs with benefits within a given program; the second allows comparison between programs.

Table 5: Sample Sheet for Cost Benefit Analysis

Benefits Estimate Worksheet **Cost Estimate Worksheet**

	Estimated Benefits:			Resources needed in Units	Estimated cost:		
Number of beneficiaries					No. of units	Unit value	Total cost
Direct benefits			Direct costs				
1.			Labor	Hours:			
2.			1.				
3.			2.				
4.			3.				
5.			Direct costs				
6.			1. Rent				
			2. Utilities				
Indirect benefits			Equipment & materials				
1.			1. Printed materials	Pieces:			
2.			2. Furnishings				
3.			3. Instructional Materials				
4.			4. Travel	Miles:			
5.			Opportunity costs				
6.			1. Child care				
			2. Food				
			3. Travel				
			INDIRECT COSTS				
Total program			Total program costs				
Benefit/ cost ratio							

STEP 7: SAMPLING FOR PROGRAM EVALUATION

Evaluation of extension programs and projects usually involves first-hand collection of data from people. The collection of data essentially involves decision about the population and a sampling plan. First, we must understand the concepts of population and sample.

Population is defined a group of individual persons, objects or items having characteristics in common, such as recipients of agricultural extension services, vegetable producers in a district, women business owners, or farm radio listeners in a province. It is the total group from which samples are taken for statistical measurement.

Rather than surveying every person in a given population, evaluators often survey a sample of the population. Why use a sample rather than a complete count? It is cheaper in terms of time, money, materials, and effort. Using statistics, combined with an appropriate sampling plan and relatively high response rate, results can be accurate and precise.

A good sample is a miniature version of the population. It is a portion, or subset, of a larger group called population, but smaller (Fink, 1995). The best sample is representative, or a model, of the population. A sample is representative of the population if important characteristics (e.g., age, educational level, ethnicity, income) are similarly distributed. Sampling involves selecting a smaller number of units from among the relevant whole group (population) in such a manner that they can be used to make estimates about the whole group.

A sample is a set of respondents selected from a larger population for the purpose of a survey. When done properly, the sample represents the characteristics of the population as a whole. Sampling saves time, money, materials and efforts without sacrificing accuracy and precision.

Sampling methods usually are categorized in two types: random (probability) sampling and purposeful (non-probability) sampling.

RANDOM (PROBABILITY) SAMPLING:

Random or probability sampling is based on random selection of units from the identified population. Random (also called probability) sampling provides a statistical basis for claiming that a sample is representative of the target population. Samples are based on random selection of units. Every member of the target population has a known probability of being included in the sample. It eliminates subjectivity in choosing a sample. It is a "fair" way of getting a sample.

Several types of random (probability) samples can be used, including the following:

- *Simple random sampling*. All individuals in the population have an equal and independent chance of being selected as a member of the sample. The list of eligible units comprising a population from which to sample is called a *sampling frame*. Members of the population are selected one at a time and independently. Once they have been selected, they are not eligible for a second chance and are not returned to the pool. One can use computer-generated lists of random numbers to select the sample. A random numbers table is sometimes used with a random starting point to identify numbered subjects.
- *Systematic random sampling*. All members in the population are placed on a list for random selection and every *nth* person is chosen after a random starting place is selected. Suppose you have a list of 4,500 households living in a watershed for which a sample of 450 is to be selected for surveying. Dividing 4500 by 450 yields 10, indicating that you have to select one of every 10 households. To systematically sample from the list, a random start is needed. You can toss a die to get a number, or consider the month of the year you were born. Suppose you were born in March,

the 3rd month of the year. This means that the 3rd name on the list is selected first, then the 13th, 23rd, 33rd, 43rd, and so on until 450 names are selected.

- *Stratified random sampling*. To ensure that certain subgroups in the population will be represented in the sample in proportion to their numbers in the population, each subgroup, called a "stratum," is separately numbered and a random sample is selected from each stratum. A clear rationale should exist for selecting any stratum. It is more complicated than simple random sampling, and using many subgroups or "strata" can lead to a large and expensive sample.
- *Cluster random sampling*. The unit of sampling is not the individual, but rather a naturally occurring group of individuals, such as a classroom, neighborhood, or club. The clusters are randomly selected and all members of the selected cluster are included in the sample. Cluster sampling is used in large-scale evaluations.

PURPOSEFUL (NON-PROBABILITY) SAMPLING

Evaluators may have to choose purposeful (non-probability) samples if accurate listings of the population are not available, resources to develop a sampling frame are limited, or obtaining co-operation from potential respondents is difficult. A purposeful sample may be chosen to be sure to include a wide variety of people based on a number of critical characteristics. Sometimes, individuals are specifically chosen to represent a certain characteristic. More frequently, evaluators choose non-probability samples because they can be conveniently assembled. A purposeful sample does not rely on random selection of units.

A purposive sample is chosen to include a wide variety of people on the basis of a number of specifically chosen and critical characteristics. Purposive sampling does not rely on random selection of units.

The following are common purposeful or non-probability samples:

- *Accidental sampling*. This is the weakest type of sample, but is the easiest to get. "Man-in-the-street" interviews are typical of accidental samples. The evaluator usually uses the first five or ten people who happen along and are willing to talk.
- *Reputational sampling*. This involves selecting specific people to respond to a survey or to be interviewed about an issue. The choice of an individual depends on someone's judgment of who is and who is not a "typical" representative of the population.
- *Convenience sampling*. A convenience sample consists of individuals who are available for data collection. For example, households living near parks or schools or persons working in a factory or business are chosen because of convenience.
- *Snowball sampling*. This type of sampling relies on previously identified members of a group to identify other members of the population. As newly identified members name others, the sample snowballs. This technique is useful when a population listing is unavailable.

Five Steps in Sampling

1. Define the population: what is its size and how varied is it?
2. Decide how much sampling error can be tolerated.
3. Determine sample size, choose sampling method, execute sampling plan.
4. Analyze data and draw conclusions based on sample information.
5. Infer conclusions back to the total population.

DETERMINING SAMPLE SIZE

Several factors need to be considered when determining sample size.

Characteristics of population: Sample size must consider the amount of variability in the population to be sampled. A relatively homogeneous population may permit a relatively small sample size. Conversely, a more heterogeneous one may require a larger population size.

Sampling error: The difference between an estimate taken from the population and that taken from the sample when the same method is used to gather the data is called the sampling error. It is larger when the sample size is small. Therefore, it is advisable to use the largest sample size possible given constraints of time, money, and materials.

Degree of precision*:* Precision measures the degree to which an estimate approximates the estimate obtained from the total population, assuming the same method of data collection was used. In designing a sample, the evaluator may begin by defining the degree of precision desired.

Margin of error: Margin of error refers to a pre-selected tolerance for amount of error in the results. What is chosen is a matter of choice, depending on the objectives of the inquiry. If we want to be relatively safe about our conclusions, then a 5 percent margin of error is acceptable. In general, more subjects are needed for a .01 alpha test than a .05 alpha test; a two-tailed test requires a larger sample size that a one-tailed test.

Confidence level: the probability that a value in the population is within a specific, numeric range when compared with the corresponding value computed for the sample. Generally, a 95 percent confidence level will give the security needed to draw conclusions for the larger population based on the sample.

Cost: Many decisions made while developing an evaluation project affect the final cost. Sample size is one of these factors. The smaller the sample size for a given set of other factors, the less the cost relative to larger sample sizes.

A frequently asked question is “How large a sample should be taken?”. As indicated above, many factors influence the size of sample. Kerlinger and Lee (2000) offer a rough-and-ready rule – “Use as large a sample as possible… the smaller the sample the larger the error, and the larger the sample the smaller the error” (p.175).

Table 6: Table for determining sample size from a given population

Population	Sample	Population	Sample		Population	Sample
10	10	220	139		1200	291
15	14	230	143		1300	296
20	19	240	147		1400	301
30	28	260	155		1600	309
40	36	280	161		1800	316
50	44	300	168		2000	322
60	51	340	180		2400	331
70	59	380	191		2800	337
80	66	420	200		3500	346
90	72	460	209		4500	353
100	79	500	217		6000	361
110	79	550	226		7000	364
120	91	600	234		8000	366
130	97	650	241		9000	368
140	102	700	248		10000	369
150	107	750	254		15000	375
160	112	800	259		20000	377
180	123	900	273		40000	380
200	131	1000	284		75000	382
210	135	1100	288		1000000	384

QUESTIONS FOR DISCUSSION AND EXERCISE

1. *Based on feedback on your evaluation plan from Chapter Four, continue planning your evaluation by completing the table below:*

Project/Program Objective(s)	**Activities Planned**	**Indicator(s) of Program Merit** (What will show the program was a success?)	**How to Collect Data to Measure Success? (What methods to use?)**	**Time and Place of Data Collection**

Nature of the Sample	**Who will Gather Evaluation Data?**	**How will Data be Analyzed?**	**Who are Your Stakeholders? How will You Share Evaluation Information with Them?**

2. *Do you plan to make use of secondary data? What kinds of secondary data are available for your evaluation?*
3. *Do you plan to gather primary data for your evaluation? What methods will you use and why?*
4. *What are the key elements of developing quality survey instruments? List them on a piece of paper and discuss with your group.*
5. *How is validity of a data collection instrument important? How do you ensure validity of an instrument?*
6. *How is reliability of a data collection instrument important? How do you ensure reliability of an instrument?*

7. *By now, you have a fairly good idea about the nature of your evaluation (quantitative, qualitative, or mixed method). Again, consider the context and objectives of the project/program you are evaluating for this course and address the following questions.*

    I. *What is your population from which you plan to gather evaluative data? Do you have more than one population? What are some characteristics of this population, e.g., educational level, sex, race, geographical spread out, etc.?*

    II. *Will you be taking a sample or will you be studying the entire population for the project/program you are evaluating?*

        i. *If you are studying the entire population, give the reasons why you have chosen to study the entire population.*

        ii. *If you are taking a sample:*

            a. *Will you have access to up-to-date sampling frame?*

            b. *What type of sample will you use?*

            c. *How large of a sample will you take?*

            d. *What method will you select to choose your sample (i.e., simple random, systematic, stratified random, stratified systematic, etc.)? Why did you choose this method of sampling?*

            e. *What will be the limitations of your sampling plan (if any)?*

CHAPTER VI

STEP 8: COLLECTING AND ANALYZING EVALUATION DATA

A sound evaluation design and sampling plan is necessary, but not sufficient, to ensure a quality evaluation. Whether you are using quantitative, qualitative, or mixed methods, collecting accurate and complete information is critical for making accurate judgments.

Quality data collection requires trained persons:

- who are familiar with agricultural extension and rural advisory service;
- who are adequately trained in evaluation methodology and data analytical tools; and
- who are very familiar with local culture and values.

Program evaluators must carefully select the data collectors. Technical and interpersonal communication skills are critical to successful data collection. Good data collectors are well organized persons who are punctual, record data clearly and accurately, and follow directions closely.

Training of data collectors is essential before actual data collection. If possible, evaluators should identify and select data collectors at the time of finalizing data collection instruments and plans. Pre-testing or pilot-testing of instruments offer an opportunity for hands-on training for data collectors.

Quality data collection also requires:

- close supervision of the data collection process;
- frequent checking for completeness of information or data collected;
- timely adjustment of procedures or tools to fit changing conditions; and
- timely data entry into a computer database for safe-keeping and analysis.

In principle, evaluations are planned when the programs or projects are planned. These plans may need to be revised and updated on a continual basis because, as the societal context and needs keep changing, such as changes in technology, road and communication networks, and market supply and demand conditions, so do extension programs. A successful extension program follows a dynamic process because it continuously adjusts its objectives and strategies to meet the needs of its clientele. Therefore, program managers need to adjust the standards and criteria for monitoring and evaluation while evaluating extension programs. In-service training and close supervision of staff reinforces their evaluation competency and helps ensure quality data collection.

Factors to consider when collecting evaluation data include:

Availability: Check if information already exists by reviewing records, reports, and census records.

Need for Training or Expert Assistance: Determine if evaluation tools/techniques require special skills on the part of data collectors and if they need to be trained in the evaluation procedure.

Protocol Needs: Make sure to acquire human subjects research approval, permission or clearance to collect information from people or other sources.

Bias: Bias means to be prejudiced in opinion or judgment. Bias can enter the evaluation process in a variety of ways. Minimize bias as much as possible by taking a random sample, using a data collection guide, assuring anonymity and confidentiality of responses, and establishing trust with subjects/respondents.

ORGANIZING EVALUATION DATA

Evaluators must develop a plan for organizing data before the data are actually collected. A good data organization plan ensures that the data will be maintained in a database that is secure and has ready access for the analysis. Statistical software (e.g., Excel, SPSS, and SAS) are available for quantitative information. NuDist and Ethnograph are commonly used qualitative analysis software.

Tips for organizing your evaluation data:

- Establish a protocol for how to receive and record the information as it comes in. Don't wait for all information to come in before recording data.
- Label all data immediately as you collect or receive it. It is particularly important to label audio/video tapes with the name of the interviewee, interviewer, and any other pertinent information.
- If questionnaires are used, record the date received, insert code number, and check off the name of respondent in the master list.
- If interview schedules are used, record date of interview, name of interviewer, and check for completeness of information.
- If data are being transcribed or transferred in some way, check to be sure that this is done accurately.

ANALYZING AND INTERPRETING DATA

Various kinds of data analysis exist for both quantitative and qualitative data. You should consider whether the analyses provide the information needed to answer the questions posed by the evaluation. Be sure that the evaluator possesses the analytical skills necessary.

QUALITATIVE DATA ANALYSIS

Qualitative data are mainly narrative data that come in many forms and from a variety of sources. They include data in the form of words or texts, pictures and expressions. Narrative data may come from:

- open-ended questions and written comments on survey questionnaires;
- personal interviews, focus group interviews, key informant interviews, case studies;
- daily journals and diaries;
- documents, photographs, reports and news articles; and testimonials, stories based on personal accounts of experience.

Qualitative analysis frequently makes use of success stories. If you plan to use a success story to illustrate the impact of your program, you should plan it well.

Key Elements of a Success Story

Title of Success Story: _____

- What was the local need? How was it determined?
- What was your role in the project? What was the role of collaborators?
- Who was your primary audience? How many did you reach and how?
- What was the outcome or result of this project, event, or activity?
- Why was this outcome or result important to helping create a sustainable agriculture community?
- Have you received any feedback from your audience about the impact of this project? Do you have any client quotes? List the feedback.
- Describe whether the results of this project or activities could benefit others in your village or state. If yes, how many farms or families in your region could benefit from this? Describe the potential impact (economic, social or environmental) of this project to your region.

Example: "Women of Hamsapur village no longer use the local money-lenders. They now borrow money from their own Mothers Group by paying 12% interest per year which is less than 1/3rd interest charged by the money-lenders. This was not the case 10 years ago when the Mothers Group did not exist in the village.

In 2005, Indragufa Community Development Foundation, a local non-governmental organization in Nepal helped organize the Mothers Group in Hamsapur. It provided an initial funding to the Mothers Group to support income generation in the community. Members of the Mothers Group could get a micro-loan with 12% interest/year to initiate an income generation project such as vegetable production, or raising a water buffalo or goat. About half of the interest goes to the local NGO who monitors the income generation project and supports other community development activities in the area. The remaining income from the interest stays within the Mothers Groups who manage their own account.

Within 5 years of operation, in 2010, the Mothers Group had supported more than 30 micro-projects to their members. The repayment rate on micro-loan was 100%. During a field visit in December 2010, members of the Mothers Group indicated, "We know each other very well. We know which member of our group needs what kind of income generation project. We know what size of micro-loan they would need. We advise them how they can maximize the use of micro-loan." They were confident that they can continue to manage micro-loans to serve the needs of their members.

Evaluators who specialize in qualitative analysis use a method called Content Analysis. Content analysis is a systematic technique for the analysis of the substance of a variety of documents. This process includes carefully reading the information, then identifying, coding, and categorizing the main themes, topics, and/or patterns in the information. Coding involves attaching some alpha-numeric symbol to phrases, sentences, or strings of words that follow a similar theme or pattern. This process allows placing these themes into a category to draw meanings.

Analysis of qualitative data is the process of bringing order to the data and organizing data into patterns, categories, and basic descriptive units. It requires reading, comprehension, and organizational skills. Interpreting qualitative data is the process of bringing meaning to the analysis, explaining patterns, and looking for relationships and linkages among descriptive dimensions. Based on interpretation, the evaluator makes judgments about the value, or worth, of a program.

Consider the following when doing qualitative analysis:

- the words used by the participants and the meaning of those words;
- the context (interpret the comments in light of the context);
- internal consistencies and inconsistencies (determine the cause of inconsistencies);
- frequency or extensiveness of comments;
- intensity of comments;
- specificity of responses; and
- dominant themes.

Taylor-Powell and Renner (2003) offer the following guidelines for qualitative data analysis:

- Get to know your data. Read and re-read the text or listen to the tapes before you begin the analysis. Quality analysis depends on understanding the data and its context.
- Review your evaluation question (what is its purpose, what questions you are going to answer from the evaluation). Focus on particular program topics and criteria.
- Classify information into different themes or patterns, and organize them into coherent categories that summarize and bring meaning to the text.
- Identify themes or patterns and connections within and between categories.
- Interpret the findings by using themes and connections to explain your findings.

INTERPRETING QUALITATIVE DATA

Data analysis focuses on organizing and reducing information and making logical or statistical inferences. Interpretation attaches meaning to information and draws conclusions. The interpretations may be influenced by the evaluator's philosophy. Therefore, not only interpretations, but also the reasons behind them should be made explicit. Useful interpretation methods include the following:

- determining whether objectives have been achieved, assessed needs have been reduced, demands met, or problems solved;
- determining the value (including non-monetary value) of a program's accomplishments;
- asking critical reference groups to review the data and to provide their judgments of successes and failures, strengths and weaknesses;
- comparing results with those reported by similar entities or endeavors; and
- interpreting results from multiple perspectives.

One method of bringing multiple perspectives to the interpretation task is to use *stakeholder meetings*. Stakeholders can be supplied in advance with the results, along with other pertinent information, such as the evaluation plan and list of questions, criteria, and standards that guided the evaluation. A meeting with stakeholders can systematically review the findings, with each participant interpreting each finding, using questions such as: What does this mean? Is it good, bad or neutral? What are the implications? What, if anything, should be done as a result of new understandings? Final evaluation reports should incorporate feedback from the stakeholders.

An example of qualitative data analysis generated from open-ended survey questions is provided in the Appendix.

Pitfalls to avoid with qualitative analysis:

- Avoid generalizing the findings across a population. Remember, qualitative analyses focus on answering, "What is unique about this individual, group, situation or issue and why?"
- Choose quotes carefully, ones that support the argument or illustrate success.
- Ensure confidentiality and anonymity of data. Do not expose or identify a respondent. Get people's permission to use their words as quotes.

QUANTITATIVE DATA ANALYSIS

Quantitative analysis involves numbers. Evaluation data usually are collected in the form of numbers, or qualitative responses, are converted into numbers using systematic coding procedures. Data may be grouped by different types of data, as obtained through different scales of measurement.

Scales of measurement

Scales of measurement refers to the type of variable being measured and the way it is measured. Different statistics are appropriate for different scales of measurement. Scales of measurement include:

Nominal: mutually exclusive and logically exhaustive categories.

Examples: marital status, gender, group membership, religious affiliation

Ordinal: ranked or ordered.

Examples: letter grades, social class, attitudinal variables

Interval: ranked and ordered in standard units of measurement.

Examples: years of age, degree, calendar year, scores on a test.

Ratio: an interval scale with an absolute zero starting point.

Examples: years of age, years of education, time, length, weight.

Understanding scales of measurement allows evaluators to use the appropriate statistical test to analyze data and report fairly and concisely the evaluation results.

DESCRIPTIVE STATISTICS TO ANALYZE EVALUATIVE DATA

Analyzing Descriptive Data: Measures of Central Tendency

The purpose of central tendency is to report a single summary score or category that best describes a set of observations. Mean, median and mode are the most common measurements of central tendency and are used to compare one group with another, identify some behavior that is unknown, or compare a group to a standard.

The mean is used for data collected at interval and ratio levels. It is the arithmetic average of all observations. You calculate mean by totaling all observations (scores or responses) and dividing by the number of observations. The mean is sensitive to “outliers” or extreme values in the observations. When your data has a few extremely small or large observations, the data are “skewed.”

Example: Twenty (20) young farmers participated in a three-week training program on improved methods of rice production. Extension educators collected pre-training rice production data on each farmer participant. After one year, each participant reported their rice yield (post-training). Fifteen farmers (response rate = 75%) reported rice yield (in Kg) decreases/increases per hectare as:

100, 30, 75, 300, 400, 300, 500, 400, 200, 100, 700, 500, 400, 500, and 200.

All participants reported an increase in rice yield as a result of participating in the training program. The minimum increase in yield was 30Kg, maximum was 700Kg, and the range was 670Kg.

The mean yield increase per Ha ($\sum X$/n) was:

(30)+(75)+(300)+(400)+(300)+(500)+(400)+(200)+(100)+(700)+(500)+(400)+(500)+(200)/15 = **314 Kg**

The median is the middle observation. It is the "in the middle value"—i.e., half of the observations are higher/larger and half are lower/smaller. It is most appropriate for ordinal variables. The median is not as sensitive to the outliers as the mean.

Example:

Observation 1: 6, 8, 13, 18, 25. N is odd. The median is 13, because half the scores fall above this number and half fall below.

Observation 2: 1, 4, 7, 8, 10, 11, 21, 22. N is even. The median is determined by summing the middle two numbers, i.e., 8 and 10, then dividing by 2. The median is 9.

The mode is used for nominal variables. It is the observation or category that occurs most frequently. The mode can be used to show the most "popular" observation or value.

Example:

Distribution A: 5, **7, 7, 7**, 8, 9, 9, 10, 11, 11

Distribution B: 4, 5, 5, 6, **7, 7**, 8, 9, 10, 11

Distribution A is unimodal, or has a single mode of 7, with 3 responses.

Distribution B is bimodal, or has two modes, 5 and 7, with 2 responses each.

When to Use Mean, Median, or Mode?

Use the mean when:

- the distribution is approximately symmetrical; or
- you are interested in numerical values.

Use the median when:

- you are interested in the typical score;
- the distribution is skewed; or
- you have ordinal data.

Use the mode when:

- the distribution has two or more peaks; or
- you want to identify the prevailing view, characteristic, or dominant quality.

ANALYZING DESCRIPTIVE DATA: MEASURES OF VARIABILITY

Variability indicates the spread or dispersion of the data. The measures of variability include range, variance and standard deviation.

Range is the difference between the largest and smallest scores in a distribution.

Example:

Given the scores of 3, 6, 8, 10, 14, and 17:

The range is 14 points because the scores range from 3 to 17.

Applied Example:

Using the example of rice yield after training:

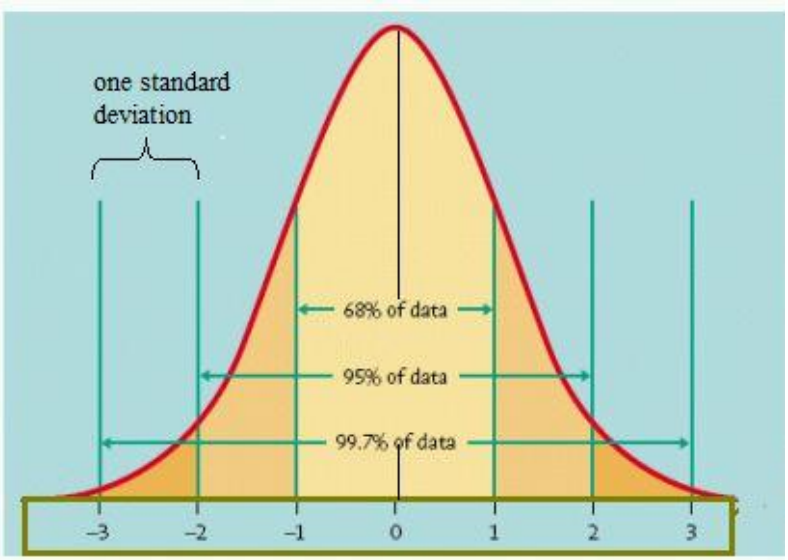
The lowest increase reported was 30 Kg/ha, the highest yield increase was 700 Kg/hectare, resulting in a range of 670 Kg.

Variance is the mean of the squares of the deviation scores. Calculate the difference (deviation) between each score and the mean of the scores, square the deviations, sum the squares and divide the sum by the number of scores minus 1.

Standard deviation is an indication of the variability of scores in a population. It measures the spread of data about their mean and is an essential part of any statistical test. It is calculated by taking the square root of the variance. This transforms variance into the same unit of measurement as the raw scores. Standard deviation is expressed in terms of "one standard deviation above the mean," or the like.

The theory of normal distribution helps us to understand and interpret standard deviation values. The main characteristics of normal curves are unimodality (one curve), symmetry (one side is the same size as the other), and certain mathematical properties (mean, mode and median values) are equal.

Figure 13: Area under standard normal distribution



Source: http://faculty.virginia.edu/PullenLab/WJIIIDRBModule/WJIIIDRBModule7.html

Example: Mean test score is 63 with a standard deviation of 11. Then one standard deviation above the mean is 74, two standard deviations is 85 and so forth. The value of this figure becomes apparent when we understand the relationship between standard deviations and percentiles in a normal curve. The area contained within +1 and -1 standard deviations of the mean includes approximately 68 percent of all scores on the distribution. Therefore, in our example, 68 percent of all scores were between 52 and 74.

Another way of assessing the meaning of the standard deviation is to compare scores with percentiles. It is known that, in a normal distribution, 96 percent of the cases are within two standard deviations from the mean. Therefore, when a raw score for one case is found to be two standard deviations above the mean, we know that the case scored higher than 96 percent of all other cases.

TESTS OF RELATIONSHIP OR ASSOCIATION

Relationship or Association: There is a relationship (or association) between variables when knowledge of one property (characteristic) of a case reduces uncertainty about another property (characteristic) of the case. A relationship (association) between variables means that variables tend to "go together" in a systematic way.

Correlation statistics measure the relationship between two variables, often between a dependent variable and an independent variable (e.g., number of years of farming and farm income), and are reported within a range of +1 (perfect positive correlation) to -1 (perfect negative correlation). A correlation coefficient value of 0 means there is no liner relationship between the variables.

We use correlations with questions such as:

Do eating habits correlate with the annual income?

Do farmers who attend extension workshops on a regular basis adopt more new practices?

Is sex associated with intention to adopt hybrid maize seed?

It is important to note that correlations identify relationships between variables, but they *do not establish causation.*

MEASURES OF ASSOCIATION IN CONTINGENCY TABLES[1]

Cross tabulation and Chi-square test are used to explore associations between variables of interest. Following is a way to group measures of association based on the scale of measurement used for variables under study.

a) Association between two nominal variables

Phi Coefficient (ϕ) (2 x 2 Contingency Table)

The phi coefficient is a useful and simple way of answering a number of evaluation or research questions about the relationship between two dichotomous variables (both nominal type, e.g. sex and adoption of IPM practices with a yes/no answer).

> **Interpretation of Results:**
>
> A Phi coefficient of zero indicates independence (no association) between variables. A Phi coefficient of +1 indicates complete dependence (association) between the variables. When there is a perfect negative relationship, Phi is -1.

[1] Adapted from Miller (1998)

b) Cramer's V (R x C Contingency Table)

Similar to Phi Coefficient, it measures association or relationship between two categorical variables with 3 or more categories, (e.g., race [white, African American and Asian American] and adoption of IPM practices [with yes/no answer]).

Interpretation of Results:

Cramer's V value lies between 0 (reflecting complete independence) and 1 (indicating complete dependence or association) between the variables.

c) Association between two ordinal variables (Ordered Categories)

Useful in examining the association between two ordinal variables (e.g., a farmer's ranking [# 1 or best farmer, # 2, # 3 and so on] in rice production and ranking in corn production in a village).

Each pair of cases is checked to determine if the relative ordering on the first variable is the same as (concordant) or reversed (discordant) from the relative ordering on the second variable.

Kendall's tau-b and tau-c are used to determine the relationship.

Tau-b is appropriate for square contingency tables, that is, when the number of rows equals number of columns.

Tau-c is appropriate for rectangular tables, that is, when the number of rows is not equal to the number of columns.

Interpretation of Results:

Tau value will be between -1.0 and +1.0. If the pairs of cases tend to be ordered the same way on both variables, a positive relationship or association between the two variables is indicated; if the pairs of cases tend to be ordered reversed on the two variables, a negative relationship is indicated; if there is no pattern in the way cases are ordered on the two variables, the variables are independent (no relationship).

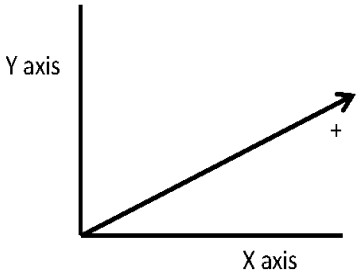
Correlation is defined as how strongly the values of an independent variable (*x)* are related to the values of a dependent variable (*y,)* with which it is paired. The word "correlation" is a derivative of the word co-relation, which helps in understanding the meaning of the term.

Relationships can be positive or negative and can be linear or non-linear.

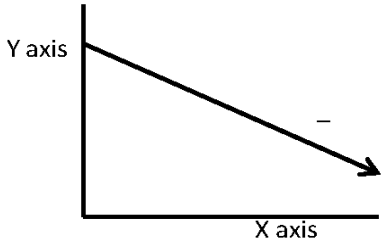
Positive relationships indicate that, as the value of *x* increases, the value of *y* increases.

Negative relationships indicate that, as the value of *x* increases, the value of *y* decreases.

Linear relationships include those in which the changes in *x* and *y* values are consistent. Non-linear relationships are not consistent in the same way. Generally, non-linear relationships indicate that as *x* increases to a point, *y* increases; after that point, y decreases. Of course, the opposite is also true.



Positive or Direct Relationship



Negative or Inverse Relationship

Pearson's Correlation Coefficient (*r*)

Pearson's correlation coefficient is based upon the deviation of points where *x* and *y* values intersect from the line of best fit. Pearson's *r* assumes two variables that are continuous and linearly related.

As an example, a Pearson *r* could be applied to height and weight (continuous) data taken from 40 sample subjects. Note that since Pearson's *r* is a measure of relationship; both the height and weight data must be collected from the identical 40 subjects (i.e., there is only one group of subjects and the data is "paired"). A resulting "*r*" value and an associated significance (probability) level would assess both the direction (+ve or direct; -ve or inverse), and the strength (between 0 and 1.00) of the relationship between the two variables.

Data should meet parametric assumptions such as normal distribution of population, and use of a random sample to gather information or data. A minimum of 30 subjects, with data on both variables for each subject, is needed to produce meaningful results.

Interpretation of Correlation Coefficients

Users of research have responded to the need to be able to discuss and interpret coefficients beyond mere numbers. Several have devised conventional terms to help express the strength of associations.

A commonly accepted set of descriptors has been (Davis, 1971):

Coefficient	Description
.70 or higher	Very strong association
.50 to .69	Substantial association
.30 to .49	Moderate association
.10 to .29	Low association
.01 to .09	Negligible association

Another commonly used set is offered by Rowntree (1981):

0.0 - 0.2	very weak, negligible
.2 - 0.4	weak, low
.4 - .7	moderate
.7 - .9	strong, high, marked
.9 - 1.0	very strong, very high

Interpretation of correlations should be done cautiously. First, a large sample size can easily produce a statistically significant *r* value, but still have very little actual strength of relationship between the variables. Secondly, correlation is not causation. As an example, player height and points scored in basketball may be significantly and positively correlated (i.e., taller players score more points), but a player scoring thousands of points won't grow an inch due to scoring ability! Neither will height always cause high scoring.

Other authors use similar descriptors. The key in using such terms is to identify the source, include it in your citation, and be consistent.

Table 7: Selection Guide for Common Statistical Methods

Scale of Measurement of Data	Statistical Method	Differences (between groups)	Testing for: Relationships (within one group)
Categorical: **Nominal** **Ordinal**	Non-parametric	Use cross-tab or Chi-square	Use contingency coefficient: • Phi coefficient • Cramer's statistic • Kendall Tau b (square table) • Kendall Tau c (rectangular table)
Nominal **Interval** **Ratio**	Parametric	Independent variable should be categorical and dependent variable should be at interval or ration scale. • Use T-test for independent groups to compare means for 2 groups • Use T-test for matched-pair to compare pre-post/before-after mean scores. • Use ANOVA to compare means for 3 or more groups.	• Use Pearson correlation to determine linear relationship between two variables measured at interval or ratio level. • Use regression to determine relationship between two or more variables. • Use discriminate analysis if independent variable is measured at interval/ratio scale and dependent variable is dichotomous.

TESTS FOR DIFFERENCES

Evaluation work usually involves making comparisons. When we evaluate, we examine whether the need gap – "what is and what ought to be" – has been reduced. In other words, we compare a program's outcome or impact data with baseline data. In some situations, we compare the data from the treatment group with data from the control group. For example, do farmers now produce more wheat per hectare than 5 years ago? Do farmers receiving extension service achieve higher crop yields than those who do not receive extension service help? When we compare, we test for differences. Different tests are used in different circumstances.

Chi-Square (x^2) is the most popular of all non-parametric inferential statistical methods. Chi-square tests for differences between categorical variables (e.g., nominal or ordinal data). Chi-square is a non-parametric statistic and, as such, requires no parametric data assumptions. The data must be categorical in nature.

T-test: A T-test is used to test the difference between two means, even when the sample sizes are small. The significance of the t statistic depends upon the hypothesis the evaluator plans to test. If you are interested in determining whether or not there is a difference between two means, but you do not know which of the means is greater, use the two-tailed test. If you are interested in testing the specific hypothesis that one mean is greater than the other, use the one-tailed test. Note that data should satisfy parametric assumptions:

- The sample is selected from populations that are normally distributed.
- There is homogeneity of variance (i.e., the spread of the dependent variable within the group tested must be statistically equal).
- The dependent variable must contain interval or ratio data.

T-Test for matched pairs: If both groups of data are contained in each data record, the appropriate t-test is for matched pairs. An example of an appropriate use of the t-test for matched pairs might be to compare pre- and post-test scores, such that each person took a pre-test (variable 1) and a post-test (variable 2). Both values are contained in each data record.

T-Test for independent groups: If each case in the data file is to be assigned to one group or the other based on another variable, use the t-test for independent groups. For example, to compare rice yield between farmers from the extension village and the control village, split the rice yields into two groups, depending on whether the person is from the extension village or the control village (each record in the date file is assigned to one group or the other).

> **Degrees of Freedom:** The degrees of freedom (abbreviation is d.f.) reflect sample size. When two independent samples are being considered, d.f. are equal to the sum of two sample sizes minus 2; i.e., d.f. = (n1 + n2 –2).

Analysis of Variance (ANOVA): ANOVA is a collection of statistical procedures for comparing the average performance of two or more groups over time. ANOVA is used to test hypotheses about two or more population means. To use ANOVA, we should gather data using random samples from normal populations having the same variance.

When two means are to be compared, the t-test is appropriate; however, if more than two means are to be compared, running several t-tests among the group means is inappropriate. When three or more means are to be compared, ANOVA can be used to test the omnibus null hypothesis, Ho = $M_1 = M_2 = ... M^j$. For example, ANOVA can answer questions such as, “Do participants from rural, sub-urban, and urban residences have different attitudes about pesticides?”

In ANOVA, only one dependent variable can be analyzed at a time, although there may be several independent variables. If there is just one independent variable, the analysis is called a one-way ANOVA; if there are two independent variables, a two-way ANOVA.

Because of its mathematical structure, ANOVA cannot prove directly that there are differences between groups. It can only prove that the opposite (that there are no differences or that the groups are the same) is not true. Thus, ANOVA tests hypotheses about the sameness or equality of behavior, which is called the null hypothesis.

QUESTIONS FOR DISCUSSION AND EXERCISE

1. *Refer to the Questionnaire in Appendix B. Critique the questionnaire using the following criteria:*
    a. *Format of the questions*
    b. *Question structure*
    c. *Question wording and clarity*
    d. *Opportunity for additional feedback*
    e. *Length of questionnaire*
    f. *Layout of questionnaire*
    g. *What suggestions do you have for its improvement?*

2. *You have already identified the evaluation project you will be using for this class. Now you must decide what type of data collection method(s) you will be using for your project (survey, interview, focus group, etc.).*
   Develop a data collection instrument for your evaluation project including a cover letter or consent form (for personal interviews). Share your data collection instrument with the group.

3. *Refer to the questionnaire in Appendix B.*

    a. *Identify the questions that yield qualitative data.*
    b. *Identity the level of measurement for questions which seek quantitative data.*
    c. *Develop a data entry scheme (codebook) for this questionnaire.*
    d. *Administer the questionnaire to about 20 people. Enter the data into SPSS or Excel. Save the file.*
    e. *Develop dummy tables for presenting findings for this study.*
    f. *Make frequency distribution tables for the variables in Question 1, 2, 3 and 4. What can you tell from the results?*
    g. *Prepare a cross-tab table for variables "interest in taking college courses in Question 2" and "gender in Question 11" with row and column percentages. Interpret the findings.*
    h. *Create a cross-tab table for "educational level" and "employment status" and interpret your findings.*

CHAPTER VII

STEP 9: COMMUNICATE EVALUATION FINDINGS

The most challenging task for evaluators is to develop useful results from the data and share the results with its users.

Program administrators and managers have a responsibility to report evaluation findings to stakeholders and other audiences who may have an interest in the results. Communication with stakeholders should occur throughout the evaluation process to help ensure meaningful, acceptable, and useful results.

Good evaluations contribute to improvements in programs and policies. Evaluators who are committed to having their work is used ensure that their findings reach intended users in a timely manner. Use is different from reporting and dissemination (Patton (1997). Use of findings means making decisions based on evaluation (e.g., improving a program by increasing funding, altering procedures, training staff, or changing policies).

Evaluation results and recommendations that are effectively communicated to stakeholders are more likely to be used than when not shared. Developing a reporting plan with stakeholders can help clarify how, when and to whom findings should be disseminated.

Who are the intended audiences? What information will be needed?

When will information be needed? What reporting format is preferred?

Reporting results: A variety of reporting procedures may be used including:

Verbal reports	*Audio-visuals*
Media appearance	*Journal or newspaper articles*
Executive summary	*News release*
Public meeting or workshop	*Newsletters, bulletins, and brochures*
Personal discussions	*Poster sessions*
Webpages	*Annual report*

Written evaluation reports usually follow a standard format containing the following key elements:

- Title page
- Table of contents
- Executive summary
- Background (of the program, its setting, and other contextual factors)
- Purpose of the evaluation
- Methods and procedures used
- Results or findings
- Conclusions and recommendations
- References
- Appendices

Reporting Tips

- Reports that are short, concise and to the point are the ones that get attention.
- Craft the style and content of the evaluation report to fit the intended audience.
- Avoid technical terms that your audience may not know.
- Use a conversational tone appropriate for the audience.
- Use a combination of long and short sentences.
- Read report aloud to check for confusing ideas and sentences.
- Write in an active voice.
- Use a logical structure for your documents.
- Allow sufficient time for writing drafts, getting feedback, proofreading, and editing.

REPORTING NEGATIVE FINDINGS

At times you may be called on to report negative findings – the program may not have met its objectives, the program is being mismanaged, or changes are needed. Evaluation can identify negative results and their potential causes or contributing factors. Reporting these difficulties can help avoid future mistakes and suggest ways to improve. However, negative findings must be reported in a manner that helps promote learning and improvement, rather than feelings of failure.

Negative findings should be reported in a manner that:

- is sensitive to the feelings of stakeholders;
- presents positive findings first;
- uses positive terms such as “accomplishments,” “in progress,” “things to work on;”
- creates an atmosphere of reflection, dialogue, and positive thinking;
- helps stakeholders think of themselves as problem solvers;
- communicates with stakeholders throughout the evaluation process; and
- helps stakeholders reflect and process negative findings.

STEP 10: USE OF EVALUATION RESULTS

Conducting evaluation does not assure that results will be used for program improvement. Many factors influence the use of evaluation results. An evaluation report can be one of the sources of information. Decision makers also can get information about an extension program from advisors, colleagues, farmer organizations, interest groups, or the media. Some decision makers may not be interested in programmatic changes if their tenure in extension is concluding. Others may not implement recommendations because of ideological or political reasons. Often, implementing recommendations requires more resources than are available.

It is important, early in the evaluation process, to solicit feedback from the primary report audiences to make sure the evaluation report will meet their needs. Evaluation experts suggest the following strategies to ensure use:

- Primary users of evaluation (stakeholders) are identified early in the process.
- Evaluation team stays in contact with primary users throughout the evaluation process.
- Potential barriers to use of evaluation results are identified and discussed with primary users.
- Preliminary results are shared with primary users.
- Primary users are involved in helping generate recommendations.
- High priority evaluation questions are adequately addressed by the evaluation.

An evaluation should not be considered complete until the findings of the evaluation are applied:

- to make decisions about a program's improvement and/or continuation; and/or
- to plan future programs.

REFLECTING ON THE EVALUATION

Professional evaluators take time to reflect on the evaluation project. This reflection differs from the actual process of the evaluation itself. Lessons and meanings from the experience should guide future evaluative studies. The process of reflection may consider the following critical questions:

Conceptual clarity: Was the evaluation well focused and were the purpose, role, and general approach clearly stated?

Description of program/project to be evaluated: Did the evaluation contain a thorough, detailed description of what was evaluated?

Recognition and representation of legitimate audiences: Did all stakeholders had a voice in developing the study and an opportunity to review results?

Sensitivity to political problems in evaluation: Was the evaluation sensitive to and coped satisfactorily with potentially disruptive political, interpersonal, and ethical issues?

Comprehensiveness & inclusiveness of data: Did the evaluation collect useful data on all important variables and issues?

Technical adequacy: Did the evaluation design and procedures yield information that meets scientific criteria of validity, reliability, and objectivity?

Appropriate methods and analysis: Were the appropriate methods chosen for data collection? Were they used correctly? Were data analyzed and interpreted carefully?

Explicit standards and criteria for judging: Did evaluation use explicit criteria and standards to make judgments?

Judgments and/or recommendations made by evaluation: Did the evaluation offer judgments and recommendations suggested by the data?

Reports tailored to audiences: Were the evaluation reports timely and appropriate for the audience?

QUESTIONS FOR DISCUSSION

1. *What criteria would you consider when developing an evaluation report?*
2. *What are the key elements of a written evaluation report? List the elements and briefly describe what should be included under each element.*
3. *Evaluation critics note the evaluation reports are not read widely. How can we improve the readership of these reports?*
4. *What would be an ideal evaluation report for people with a lower level of education and/or training in developing countries? How would you communicate results with these stakeholders?*
5. *How would you ensure the use of evaluation results to improve future agricultural extension programs and activities?*

APPENDIX A: ONLINE RESOURCES ON PROGRAM EVALUATION

Frechtling, Joy (2002). **The 2002 User-Friendly Handbook for Project Evaluation**. Division of Research, Evaluation and Communication, National Science Foundation.
www.nsf.gov/pubs/2002/nsf02057/start.htm

Michigan State University – Extension: Evaluation Resources
www.msu.edu/~suvedi/Resources/Evaluation%20Resources.htm

Online Evaluation Resource Library. http://oerl.sri.com/

Pennsylvania State University - Extension
http://extension.psu.edu/evaluation

University of Wisconsin - Extension
Program Development and Evaluation. www.uwex.edu/ces/pdande/evaluation/evaldocs.html

W.K. Kellogg Evaluation Handbook (2000). W.K. Kellogg Foundation, Battle Creek, MI.
www.wkkf.org/documents/wkkf/evaluationhandbook/

APPENDIX B: SAMPLE COVER LETTER TO A MAIL SURVEY

Dear Residents:

We are conducting a district wide survey of citizens to assess your higher education and training needs, how you have been utilizing services of our Regional education Center, and determine what types of higher educational programs or vocational training you would like to see for citizens in this area.

Why you? A random sample of citizens in the district mailing list included your name. Will you help us out? We estimate that the survey will take about 15 minutes to complete. Knowing that you are busy, we designed the format to make it easy for you to respond quickly. Your response to the survey can help us develop higher educational and vocational training programs to better meet your needs.

Your completion of this survey is completely voluntary. You are free not to answer any questions or to stop participating at any time. All responses will be kept confidential by the researchers to the maximum extent allowable by law. There are no risks or individual benefits associated with completing this survey. Our reports will not associate any responses with any individual respondents unless express written permission to do so has been secured. If you have any questions regarding your rights as a participant, please contact ____ _____, Director of Human Research Protection Programs at Michigan State University. Phone number is _____; email address is _____.

By completing the survey you indicate your voluntary consent to participate in this study and have your answers included in the educational needs assessment data set. Your response will be treated as confidential and no individual responses will be identified. Once you have completed the questionnaire, fold it and return it to us in the enclosed pre-stamped, pre-addressed envelope.

We appreciate your cooperation and thank you in advance for your time and assistance in completing this important questionnaire. As a token of appreciation, enclosed is a ball pen for you. If you have questions about this survey please contact ___ ____ at ___ ___ ____. Thank you very much for your help!

APPENDIX C: SAMPLE QUESTIONNAIRE

Higher Education Needs Public Opinion Survey

Background

The ___ ____ Higher Education Task Force is studying the needs for additional higher education services within this area. This survey is designed to uncover citizens' higher educational needs and identify gaps in services to meet these needs. Findings of this study will help to decide additional higher education offerings at the Regional Education Center.

Please complete this survey and return it in the enclosed self-addressed stamped envelope.

Survey Questions

1. Have you or members of your family ever taken classes at the ___ Education Center?
_____ Yes _____ No

If YES, what type of educational programs did you attend? [check all that apply]

a. Credit course(s) for college
b. Non-credit course(s)
c. Certificate course(s)
d. Lifelong education course(s)
e. Other (specify)_____

2. How interested would you or your family members be in taking college or continuing education courses in Gaylord if they were offered?

_____	_____	_____	_____	_____
No Interest	Low interest	Moderate interest	High interest	Very high

3. What type of programs should be offered at the Regional Education Center?

Educational Programs	**I'm personally interested**	**I'm not interested but should be offered**
a. 2 yr. Associate degree	_____	_____
b. 4 yr. Bachelor's degree	_____	_____
c. Graduate level courses	_____	_____
d. Licensing and certification programs	_____	_____
e. Apprenticeship programs	_____	_____
f. Supplemental courses	_____	_____
g. Lifelong education classes	_____	_____

5. In which of the following skill areas would you or your family members be interested in as topics for lifelong education and/or college courses? *Please check the appropriate column of interest and circle "C" for credit or "NC" for non-credit to indicate* your interest.

	I'm personally interested	**I'm not interested but should be offered**	**Credit / Non credit**
a. Communication skills: Reading, writing, grammar, listening and telephone skills.	_____	_____	C/NC
b. Computer and technology: Excel, Lotus, MS Word, Word Perfect, e-mail, etc.	_____	_____	C/NC
c. Manufacturing skills: Tool making, using calipers, blueprint reading and machining	_____	_____	C/NC
d. Interpersonal skills: Team building, leadership, team work, customer service, listening skills	_____	_____	C/NC
e. Math/accounting skills: General applied math such as book keeping, budgeting	_____	_____	C/NC
f. Work ethic: Attitude, desire to work and learn, pride in work quality, punctuality, self motivation, etc.	_____	_____	C/NC

5. What type of education delivery method do you prefer?
   a. Classroom
   b. Satellite or I.T.V. classroom
   c. Online/Internet
   d. Other (specify)_____

6. What season of the year would best serve your needs in terms of attending a college class?
   ____ Summer _____ Fall _____ Winter ____ Spring

7. What time of day would best serve your needs in terms of attending a college class?
   Week days (Monday-Friday) ____ Morning _____Afternoon _____ Evening
   Weekends (Saturday-Sunday) ____ Morning _____Afternoon

8. What services do you need to help you attend college or continuing education classes?*(Check all that apply)*

a. Career counseling
b. Student financial aid
c. Remedial education
d. Release time from work
e. Child care programs
f. Credit transfer service
g. Transportation
h. College Level Examination Program

9. Would you be willing to financially support the development of a higher education facility in Otsego County? ___ Yes ___ No ___Undecided

If YES, how would you like to support the higher education facility?

a. Approve a millage increase (How much millage increase would you support? ____)
b. Make a personal donation/gift (How much? _____)
c. Assist in finding other sources of support like Foundations, Corporate sources, etc.
d. Other (specify) _____

10. What is your age group?

a. Younger than 25 years
b. 25-34 years
c. 35-44 years
d. 45-54 years
e. 55-64 years
f. 65 years and older

11. Your gender: ___Male ___Female

12. What is the highest level of formal education you have completed?(check only one)

a. Some high school
b. High school diploma or equivalent
c. Some college
d. Technical or trade school certification
e. 2- year associate degree
f. 4 year college degree
g. Some graduate work
h. Graduate degree

13. What is your employment status? (Check only one)

a. Work full time
b. Work part time, looking for full time work
c. Work part time, not looking for full time work
d. Unemployed, looking for work
e. Unemployed, not looking for work
f. Student
g. Retired

14. Besides yourself, how many members in your family might be interested in attending college or continuing education courses at the Education Center, if they were to be offered? _____ persons

15. Please use this space to comment on any things about higher education in your village or district:

_____

_____

Thank you for your cooperation.

APPENDIX D: SAMPLE SHORT COURSE EVALUATION FORM

Educational Event: _____ Date: _____

Here is your chance to provide feedback. We need it to plan and improve future educational programs and events. Please take a minute to share your views with us.

2. Please describe specific skills, knowledge, ideas, etc. that you gained from this course:

3. Please describe briefly how you intend to apply the knowledge and skills gained from this course on your farm/in your work.

Please rate this educational program/event by checking the choice that best describes your impression:

4. The Course/Educational Event

How familiar were you with the content of this course?	__ Very familiar	__ Some	__ All New
How would you rate the class notes (clarity, accuracy)?	__ Well written	__O.K.	__ Poorly written
Time allotted for this class was:	__ Too much	__ Just right	__ Too little

5. The Instructor:

Instructor's knowledge of the subject matter:	__ Poor	__ Fair	__Good	__ Excellent
Instructors' teaching/communication skills:	__ Poor	__ Fair	__Good	__ Excellent

6. The Learning Facilities:

Materials and equipment used were:	__ Poor	__ Fair	__Good	__ Excellent
Teaching learning environment was:	__ Poor	__ Fair	__Good	__ Excellent

6. Perceptions of the Program/ Event:

(a) I will be able to apply the information/skills.	__ Agree	__Undecided	__ Disagree
(b) I have improved my knowledge and skills.	__ Agree	__Undecided	__ Disagree
(c) I learned things that can improve my farming	__ Agree	__Undecided	__ Disagree
(d) I will recommend this class/event to others.	__ Agree	__Undecided	__ Disagree

7. Suggested topics for future training sessions or events I would be interested in attending:

_____

_____

Thank you.

APPENDIX E: SAMPLE INTRODUCTION TO FOCUS GROUP

Good evening and welcome to our focus group session. Thank you for taking the time to join our discussion on agricultural extension services in this village. My name is _____ and I represent _____.

Assisting me is _____ from _____. We want to learn about your experiences with and opinions about agricultural extension services for this community. We have invited people who live in several parts of this village to share their ideas.

You were selected because you have certain things in common and have experiences that are of particular interest to us. You are farmers or know the problems farmers face in this village. We are particularly interested in your views because you are representative of others in the village.

Today we will be discussing agricultural extension programs in this community. These include all the ways you gain new information about farming and agribusiness. There are no right or wrong answers, but rather differing points of view. Please feel free to share your point of view even if it differs from what others have said.

Before we begin, let me remind you of some ground rules. Please speak up, but only one person should talk at a time. We are recording the session because we don't want to miss any of your comments. If several are talking at the same time, the recording will get garbled and we'll miss your comments. We will be on a first-name basis tonight, and in our later reports, there will not be any names attached to comments. You may be assured of complete confidentiality of your responses. Keep in mind that we're just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.

Our session will last about an hour and we will not be taking a formal break. Well, let's begin.

APPENDIX F: SAMPLE FOCUS GROUP QUESTIONS

Agricultural Extension Service Improvement Project

Please tell us your first name and briefly describe where you live and what you do.

Questions for focus group session:

1. In your opinion, what makes a farmer successful?
2. What things in your communities help make farmers more successful?
3. Think about your community. What are the greatest concerns or needs of the farmers and agribusiness operations in your community?
4. Let us follow-up on question 3 about the greatest educational or informational needs of farmers and agribusinesses. What needs are being addressed satisfactorily by public and private agencies serving farmers? What needs are not addressed well?
5. What do people say about the agencies serving farmers in this village or district? Are they working together? What can be done to improve their services?
6. How can citizens like yourself help to improve agricultural services of these organizations in your community?
7. How can the government improve its services to farmers and agribusiness operators in this area?
8. Have we missed anything?

Thank you for your time and input. Please feel free to call the extension office if you have questions or additional suggestions.

APPENDIX G: SAMPLE CASE STUDY

The Michigan Sugarbeet Advancement (SBA) program was initiated in 1997 to help sugar beet farmers adapt to economic and environmental changes through research and dissemination of educational information. In 2006, an evaluative study was conducted to understand SBA's impacts on Michigan's sugar beet industry. It attempted to understand (a) the credibility of SBA in beet research and educational information, (b) SBA's dissemination of research-based information to growers, and (c) changes in sugar beet production practices by the growers due to this information. The study also attempted to learn about sugar beet grower concerns and specific suggestions from growers for educational program offerings.

After a careful review of SBA activities and discussion with SBA Extension Educator and affiliated growers, a survey instrument utilizing traditional mailing methods was developed and delivered to Michigan sugar beet farmers. The population included 1,342 sugar beet growers across Michigan. Following are some of the key findings of this evaluation of the SBA program.

- Respondents came from various Michigan counties, with the majority farming in Huron, Saginaw, Tuscola and Sanilac.
- The average sugar beet farmer cultivates about 1,300 acres, of which about 250 acres are devoted to sugar beet production.
- Farmers averaged 50 years of age and worked on family-owned farms full time.
- About three-quarters of survey respondents plan to give their farms to family members when they retire.
- Over three-quarters of respondents have access to a computer, the Internet and e-mail.
- The majority of growers indicated that SBA is their preferred source of information. Respondents indicated that SBA should take the lead in educational programming and rated SBA overall as the most heavily relied on source for research-based information.
- Over two-thirds of participants participated in, attended, or used SBA's farm meeting/workshops, the Bean and Beet Symposium, Sugar Beet Seed Week, "On-farm Research Demonstration" SBA publications, information from quarterly newsletters, Cercospora Leafspot bulletin and production tip cards (tips for maximizing sucrose production).
- Conversely, two-thirds or more of respondents indicated that they had not attended harvester clinics, used the SBA website, or had a local Extension Educator visit their farm.
- The information provided by the SBA programs was deemed to have helped farmers make positive changes in their practices by about two-thirds of participants. Also, two-thirds of respondents indicated that SBA information had helped increase their income due to changes in production practices.
- Overall, yields have increased from 18 tons per acre in 1997 to 24 tons in 2006.
- Major grower concerns can be grouped into four categories: profit, disease control, industry stability, and MI Sugar Company stability.

APPENDIX H: EXAMPLE OF QUALITATIVE DATA GENERATED FROM OPEN-ENDED SURVEY QUESTIONS

Michigan State University Extension conducted a statewide survey in 2008 to identify the major areas of educational need for Michigan farmers and agribusiness operators. Surveys were mailed to a stratified random sample of dairy, livestock, swine, cash crop, fruit, vegetable, and nursery/greenhouse producers. Respondents indicated needs for more educational support in the areas of farm management, production skills, environmental issues, sustainable agriculture, biotechnology, and small farm management (Suvedi, Jeong, and Coombs 2010).

Category	Rate	Percent	Example of educational need
Business, Bookkeeping and Marketing Skills	81	24.3	"Economic focus on world markets." "How to figure profitability in changing economy." "Profit/cost management."
Sustainable Farming Practices	31	9.3	"Increase emphasis on no-till farming." "Soil conservation is top priority." "Carbon footprint issues."
Management and Care of Livestock and Animals	27	8.1	"Knowing how to care for and feed animals efficiently." "Livestock disease prevention."
Chemicals and Fertilizer	27	8.1	"Proper use of pesticides." "Risks with pesticides/herbicides."
Pests and Diseases	24	7.2	"Diseases and pests in crops." "Environmentally friendly weed and insect control."
Crop Production	23	6.9	"Increase crop yields." "Productivity issues-yield and effective irrigation."
Farm Management	17	5.1	"How to handle dry spells." "Help keeping costs down in producing feed."
Laws and Regulations	16	4.8	"Help with understanding and complying with new state and federal directives."
Biotechnology	14	4.2	"Biofuel for small/medium farmers." "Biofuel production."
New Technology	12	3.6	"GPS-controlled systems." "More exposure to innovation."
Organic Farming	11	3.3	"Producing organic products." "Promote organic-style farming practices."
Water	10	3.0	"Education in water re-use and water conservation." "Methods [and] water management issues."
Labor Relations and Human Resources	8	2.4	"[How to] train and educate your employees." "Managing human resources."

Environmental Issues (general)	8	2.4	"Environmental issues" "How to keep good farm standards with little environmental impact."
Financing	7	2.1	"How to get Michigan and federal loans at the same rate the larger farmers seem only to get."
Testing	6	1.8	"More detailed and independent tests." "Soil analysis."
Waste Management	6	1.8	"Manure management." "Care and disposal of toxic wastes."
Alternative Farming Practices	5	1.5	"Alternative feeds/feeding"
Total	333	100	

REFERENCES

Aker, J.C. (2010, October). *Dial "a" for agriculture: using information and communication technologies for agricultural extension in developing countries*. Retrieved from cega.berkeley.edu/materials/AfDR_Aker_Dial-A-for-Agriculture_P-S.pdf

Archer, T. M. (1988). *Edge guide to evaluation: analyzing qualitative data.* Ohio State University, Shelby County Extension .

Archer, T., & Layman, J. (1991). "Focus Group Interview" Evaluation Guide Sheet, Cooperative Extension Service, Ohio.

Ary, D., Jacobs, L.C., & Razavieh, A. (1996). Introduction to Research in Education. New York: Harcourt Bruce College Publishers.

Ary, D., Jacobs, L.C., & Razavieh, A. (2002). *Introduction to research in education*. Wadsworth Pub Co.

Axinn, G.H. United Nations, Food & Agriculture Org. (1988). *Guide on alternative extension approaches*. Rome.

Bailan, E.S. (1982). How to design, analyze, and write doctoral research. New York: University Press of America.

Bailey, L.H. (1908). *The state and the farmer*. London: Macmillan Company.

Bennett, C., & Rockwell, S.K. (1995). *Targeting outcomes of programs (TOP): An integrated approach to planning and evaluation*. Lincoln, NE: Cooperative Extension, University of Nebraska.

Bennett, C. F. (1979). Analyzing impacts of Extension programs. Washington, DC: US Department of Agriculture, Science and Education Administration (ES C-575).

Bonnen, J.T. (1998). The land-grant idea and the evolving outreach university. In R.M. Lerner and L.A. Simon. (Eds.), *University-community collaborations for the twenty-first century.* Garland Publishing, Inc.

Boyle, R, & Lemaire, D. (1999). *Building effective evaluation capacity: lessons from practice*. Transaction Pub.

Casey, R.,Andrews, M., &Werner, W. (1988). How can we do it? An evaluation training package for development educators. British Columbia, Canada: Research and Development in Global Studies.

Casely, D., & Kumar, K. (1988). *The collection, analysis and use of monitoring and evaluation data*. John Hopkins Press for The World Bank. Baltimore and London.

Christopolos, I., Sandison, P., & Chipeta, S. (2011). *Guide to extension evaluation*. Aarhus, Denmark: Printbutikken.

Clay, J. (2010). *Agriculture from 2000 to 2050- the business as usual scenario*. Retrieved from http://dels.nas.edu/resources/static-assets/banr/AnimalProductionMaterials/

Davis, J. A. (1971). Elementary survey analysis. Englewood Cliffs, NJ: Prentice-Hall.

den Biggelaar, C., &S**u**vedi, M. (2000). Farmers' definitions, goals, and bottlenecks of sustainable agriculture in the North-Central Region. *Agriculture and Human Values* **17:** 347–358.

Devore,J., & Peck, R. (1986). Statistics, the exploration and analysis of data. St. Paul, MN: West Publishing.

Dillman, D. A. (1995). "Survey Methods." Class notes in AG*SAT Graduate Course in Program Evaluation in Adult Education and Training, University of Nebraska-Lincoln.

Dillman, D. A. (2007). Mail and internet surveys: The tailored design method. Hoboken, NJ: John Willey & Sons.

Ejeta, G. (2010). African green revolution needn't be a mirage. *Science*, 327, 831-832.

Falconi, CA, 1993. Economic evaluation. In: Horton, D, Ballantyne, P, Peterson, W, Uribe, B, Gapasin, D and Sheridan, K, Editors, 1993. *Monitoring and evaluation agricultural research: a sourcebook*, CAB International, Cambridge, pp. 65–75.

Fink, A. (1995). How to sample in surveys. Thousand Oaks, California: Sage.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B.R. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*, Second Edition. New York: Longman.

Fraenkel, J. R., & Wallen, N.E. (1996). How to design and evaluate research in education. New York: McGraw-Hill Inc.

Frechtling, J., Frierson, H., Hood, S., & Hughes, G. (2002). The 2002 User-Friendly Handbook for Project Evaluation. Division of Research Evaluation and Communication, National Science Foundation. www.nsf.gov/pubs/2002/nsf02057/nsf02057_1.pdf

Gay, L.R. 1976. Educational research: Competencies for analysis and application. Columbus: Charles E. Merrill Publishing.

Gray, S.T. (Ed.). (1993). *Leadership is: a vision of evaluation*. Washington, D.C.: Independent Sector.

Hopkins, K. D., & Glass, G.V. (1978). Basic Statistics for the Behavioral Sciences._ Englewood Cliffs, NJ: Prentice-Hall.

Hummelbrunner, R. (2011). Systems thinking and evaluation. *Proceedings of the Systemic approaches in evaluation.* Eschborn: GIZ.

Khandker, S., Koolwal, G., & Samad, H. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.

Krueger, R.A., & Casey, M.A. (2000). Focus groups: A practical guide for applied research. 3rd edition, Thousand Oaks, California: Sage.

The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: how to assess evaluations of educational programs*. Sage Publications, Inc.

Kerlinger, F, & Lee, H. (2000). *Foundations of behavioral research*. Wadsworth Pub Co.

Marynowski, S., Denny, C., Colverson, P., & Hill, K. (Ed.). (2006). *Best practices guide to program evaluation for aquatic educators*. Virginia: Recreational Boating & Fishing Foundation.

McCaslin, N.L. (1995). Evaluation: an imperative for public education and training programs. *Professorial inaugural lecture series.* Ohio State University.

Milburn, L.S., Mulley, S.J., & Kline, C. (2010). The end of the beginning and the beginning of the end: The decline in public agricultural extension in Ontario. *Journal of Extension*, 48(6).

Miller, L. (1998). Appropriate analysis. *Journal of Agricultural Education*, *39*(2), 1-10.

Misra, D.C. (1998). Monitoring extension programmes and resources. In Swanson, B.; R. Bentz, & A. Sofranko. (Eds.), *Improving agricultural extension. A reference manual* Rome: Food and Agriculture Organization of the United Nations.

Mueller, D.J. (1986). Measuring social attitudes. New York: Teachers College Press.

Naik, K.C. (1968). A history of agricultural universities. Delhi: U.S. Agency for International Development.

Neito, R., &Henderson, J.L. (1995). Establishing Validity and Reliability. Ohio State Cooperative Extension.

SRI International. (n.d.) *Online evaluation resource library*. Retrieved from http://oerl.sri.com/

Patton, M. Q. (1997). Utilization-focused evaluation: The new century text. Thousand Oaks: Sage.

Percy, R. (2005). The contribution of transformative learning theory to the practice of participatory research and extension: Theoretical reflections. *Agriculture and Human Values* 22, 127–136.

Pound, B., Gundel, S., Martin, A., & Apenteng, E. (2010). Meta-evaluation of extension evaluation case studies. National Resources Institute. University of Greenwich, Kent, UK.

Pullen, P.C. (2010). Woodcock Johnson III Diagnostic Reading Battery. Retrieved from http://faculty.virginia.edu/PullenLab/WJIIIDRBModule/WJIIIDRBModule7.html.

Qamar, M.K. (2003). Agricultural extension in Asia and the pacific: time to revisit and reform. *Proceedings of the Seminar on enhancement of extensions systems in agriculture.* Faisalabad, Pakistan.

Rasmussen, W. (1989). *Taking the university to the people: seventy-five years of cooperative extension*. Iowa State University Press.

Richardson, J. G., & Moore, C.L. (2002). *Determining extension program economic benefit values*. AEE 01-2002, Department of Agricultural and Extension Education, NC State University, Raleigh, NC.

Ricker , K., Brown, L., Leeds, C., Leeds, R., Bouton, E., & Volgstadt, C. Ohio State University Extension, (1998). *Water quality project evaluation: handbook for objectives-based evaluation of water quality projects* (Bulletin 868-98)

Rivera, W., & Gustafson, D. (1991). *Agricultural extension: worldwide institutional evolution and forces for change*. Elsevier Science Ltd.

Rogers, E.M. (2003). *Diffusion of innovations*. Free Press.

Rowntree, D. (1981). Statistics without tears. New York: Charles Scribner's Sons.

Salant, P., &Dillman, D.A. (1994). How to conduct your own survey. New York, NY: John Willey & Sons, Inc.

SPSS, Inc . (2006). *SPSS 15.0 brief guide*. Prentice Hall.

Suvedi, M., Jeong, E., and Coombs, J. (2010). *Educational needs of Michigan farmers. Journal of Extension* 48 (3): 3RIB7. www.joe.org/joe/2010june/rb7.php

Suvedi, M., & Morford, S. (2003). Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia. Forrex-Forest Research Extension Partnership, Kamloops, B.C. Forrex Series 6. www.msu.edu/~suvedi/Resources/Documents/4_1_FS6.pdf

Suvedi, M., Heinze, K., & Ruonavaara, D. (2008). How to conduct evaluation of extension programs. In Suvedi, M., Singh, B., Vijayaragavan, K., Padaria, R., & Wason, M. (Eds.) *Evaluation capacity building in rural resource management: A manual*. India: Golden Printers & Publishers.

Suvedi, M., Singh, B., Vijayaragavan, K., Padaria, R., & Wason, M. (2008). *Evaluation capacity building in rural resource management: a manual*. India: Golden Printers & Publishers.

Swanson, B. E., &Claar, J.B. (1984). *Agricultural extension: a reference manual*. The history and development of agricultural extension. Food & Agriculture Org.

Swanson, B.E., Bentz, R., & Sofranko, A.J. (Ed.). (1997). *Improving agricultural extension*. Rome: Food and Agriculture Organization of the United Nations.

Swanson, B. (2008). *Global review of good agricultural extension and advisory service practices*. Food and Agriculture Organization of the United Nations, Research and Extension Division.

Swanson, B., & Rajalahti, R. The World Bank, Agriculture and Rural Development. (2010). *Strengthening agricultural extension and advisory systems: procedures for assessing, transforming, and evaluating extension systems* (Paper 45).

Taylor-Powell, E., & Renner, M. University of Wisconsin-Extension, (2003). *Analyzing qualitative data.* (G3658-12). Madison, WI: Cooperative Extension Publishing Operations.

Taylor-Powell, E., & Henert, E. University of Wisconsin-Extension, (2008). *Developing a Logic Model: Teaching and Training Guide*. Madison, WI: Board of Regents of the University of Wisconsin System. http://www.uwex.edu/ces/pdande/evaluation/pdf/lmguidecomplete.pdf.

Technical Advisory Division, Bureau for Programme Policy and Evaluation, United Nations. (1991). *Agricultural extension: programme advisory note*. New York: United Nations Development Programme.

Uribe, B., Horton, D., 1993. Logical framework. In: Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D., Sheridan, K. (Eds.), Monitoring and Evaluating Agricultural Research: A Sourcebook. CAB International, Wallingford, pp. 113–119.

USAID, Bureau for Policy, Planning, and Learning. (2011). *Evaluation Policy.* http://www.usaid.gov/evaluation/USAID_EVALUATION_POLICY.pdf?020911

USAID, Center for Development Information and Evaluation, (2000). *Performance monitoring and evaluation tips: building a results framework* (PN-ACA-947).

U.S. Government Initiative, (2011). *Feed the future: global food security research strategy.* Washington, D.C.: Retrieved from http://www.feedthefuture.gov/documents/FTF_research_strategy.pdf

Weiss, C.l H. (1998). *Evaluation (2nd Edition).* New Jersey: Prentice Hall.

Wholey, J. S.; Harty, H. P., & Newcomer, K.E. (Eds.). (1994). Handbook of practical program evaluation. San Francisco: Jossey-Bass Publishers.

Worthen, B., Sanders, J., & Fitzpatrick, J. (1997). *Program evaluation: alternative approaches and practical guidelines*. Longman Pub Group.

Yin, R. K. (1984). Case study research: design and methods. Applied Social Research Methods Series. Vol. 5. Newbury Park, California: Sage.