ARDA

AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON, D. C. 20523
BIBLIOGRAPHIC INPUT SHEET

FOR AID USE ONLY
Batch 93 ARDA D

1. SUBJECT CLASSIFICATION	A. PRIMARY: Food production and nutrition	AE30-0000-0000
	B. SECONDARY: Development	

2. TITLE AND SUBTITLE
Adaptive field-testing for rural development projects

3. AUTHOR(S)
Olson, C.V.

4. DOCUMENT DATE	5. NUMBER OF PAGES	6. ARC NUMBER
1978	78p.	ARC

7. REFERENCE ORGANIZATION NAME AND ADDRESS
Dev.Alts.

8. SUPPLEMENTARY NOTES *(Sponsoring Organization, Publishers, Availability)*

9. ABSTRACT

10. CONTROL NUMBER
PN-AAG-084

11. PRICE OF DOCUMENT

12. DESCRIPTORS
Adaptive field-testing
Farms, small
Handbooks
Project design
Project management
Rural areas

13. PROJECT NUMBER

14. CONTRACT NUMBER
AID/CM/ta-C-73-41 GTS

15. TYPE OF DOCUMENT

AID 590-1 (4-74)

ADAPTIVE FIELD-TESTING FOR RURAL DEVELOPMENT PROJECTS

By Craig V. Olson

Submitted to
Office of Rural and Administrative Development
Agency for International Development
Under Contract No. AID/ta-C-73-41

Development Alternatives, Inc.
1823 Jefferson Place, N.W.
Washington, D.C. 20036

August 15, 1978

TABLE OF CONTENTS

PREFACE. iii

CHAPTER ONE: THE WHAT AND WHY OF ADAPTIVE FIELD-TESTING . 1

WHAT IS ADAPTIVE FIELD-TESTING? 2
WHY ADAPTIVE FIELD-TESTING? 4
WHAT CAN BE TESTED? 6

CHAPTER TWO: SOCIAL EXPERIMENTATION AND ADAPTIVE FIELD-TESTING. 9

THE ELEMENTS OF SOCIAL EXPERIMENTATION. 10
THE NEED FOR ADAPTATION 17

CHAPTER THREE: DEALING WITH UNCERTAINTY: SOME EXAMPLES FROM EXPERIENCE 21

LEVELS OF UNCERTAINTY 21
EXAMPLES OF ADAPTIVE FIELD-TESTING. 26

CHAPTER FOUR: METHODOLOGICAL OPTIONS. 47

RESEARCH DESIGN 47
SAMPLING. 51
DATA COLLECTION 56

CHAPTER FIVE: INTEGRATING OPTIONS AND POSSIBILITIES INTO A RURAL DEVELOPMENT PROJECT. 61

TIME-PHASING OF EXPERIMENTS 61
ORGANIZATION OF RURAL DEVELOPMENT PROJECTS USING ADAPTIVE FIELD-TESTING. 65

CHAPTER SIX: TEN QUESTIONS. 67

INTRODUCTION. 67
QUESTIONS 67

PREFACE

This Handbook is the result of a series of contracts issued to Development Alternatives, Inc. (DAI), by the Office of Rural and Administrative Development and its immediate predecessors within the Agency for International Development. The original research contract culminated in the publication of the report, *Strategies for Small Farmer Development*, submitted to AID in 1975. Subsequent DAI involvement in the design of rural development projects generated a report entitled, *The "New Directions" Mandate: Studies in Project Design, Approval and Implementation*, first submitted to AID in 1977. In a related contract, DAI prepared a state-of-the-art study on *Information for Decisionmaking in Rural Development* for the same office.

A central theme of all these reports has been the need for flexibility in project designs. The reports have argued that project managers should enjoy the freedom to change directions and channel resources in different ways as new information is obtained during project implementation. Two principal and mutually reinforcing methods of obtaining relevant information are through project information systems and adaptive field-testing.

A principal source of the material used in this Handbook is an original essay, written by Dr. A. H. Barclay, Jr., of the DAI staff entitled "The Role of 'Experimentation' in Development Projects." This essay became Chapter Seven of the *"New Directions" Mandate* report. Other sources included the *Strategies* report, the *Information for Decisionmaking* report, standard texts in social experimentation and evaluation research, as well as the DAI experience in project development and evaluation work over the last five years. The task of putting all this material in the present Handbook form was undertaken mainly by Dr. Craig V. Olson of the DAI staff.

The result, we hope, is a Handbook that will prove useful to project designers and managers whose projects might benefit from adaptive field-testing.

Donald R. Mickelwait
President

August 15, 1978

CHAPTER ONE

THE WHAT AND WHY OF ADAPTIVE FIELD-TESTING

As a project designer, you would like to introduce a chemical fertilizer to a group of small farmers; you believe that the fertilizer will increase yields, but you are not sure whether its cost and logistics difficulties and local reaction to its introduction will make it worthwhile.

As an advisor to a new agricultural supply cooperative, you are not sure whether a credit program is necessary or whether the local farmers can purchase inputs with cash; if credit is necessary you are unsure how high to set the interest rate.

As a project manager, you are working with a PP that calls for you to introduce a complex "package" of agricultural inputs and services to a local population; you have your doubts about the cost effectiveness and local acceptability of some parts of the package.

Do these problems sound familiar? Have you ever helped design or worked in the implementation of a development project in which you were uncertain about what activities, what interventions would be most effective, or do you anticipate encountering such problems? If the answer to these questions is yes, this handbook on adaptive field-testing may prove useful to you.

The handbook is written for practitioners of development: for project designers, project managers and project technicians who have first line responsibility to bring the benefits of development to local populations, and who must make important decisions under conditions of uncertainty. In practical terms it describes how, under a variety of circumstances, adaptive field-testing may be used to reduce the uncertainty under which decisions are made.

Examples used in the handbook are primarily taken from agricultural and rural development projects. However, readers may see how many of the principles of adaptive field-testing may be used in a greater variety of development projects.

WHAT IS ADAPTIVE FIELD-TESTING?

Adaptive field-testing may be defined as a process of "experimentation" conducted within the context of an ongoing development project, which aims at predicting with greater certainty the outcome of an intervention. It is a way of testing the appropriateness or effectiveness of a single technology or intervention, or of choosing among competing technologies or interventions. (The word "experimentation" is in quotation marks because, while the process often resembles an experiment, adaptive field-testing rarely lends itself to the scientific rigor normally associated with experimentation.)

The emphasis in adaptive field-testing is on reducing uncertainty to a level where the marginal costs of additional testing no longer justify the marginal increase in certainty that would be gained by the additional testing. Thus adaptive field-testing does not aim at eliminating uncertainty altogether; even if this were possible the upfront costs and time needed would be prohibitive in most cases. Adaptive field-testing is, rather, a practical way of making better decisions while keeping decision costs low.

It should be noted that the use of adaptive field-testing in no way obviates the need for basic agricultural or other development research. Agricultural research stations have played and will continue to play a critical role in providing improved technologies to farmer populations. However, the output of such research stations should be seen as providing only a range or selection of possible interventions, rather than definitive solutions. In virtually every case these technologies must be customized to fit local ecological and social conditions; and it is in this customizing process that adaptive field-testing has a role to play. Furthermore adaptive field-testing delivers results and benefits as well as learning. Thus it is frequently more palatable to host country governments or donor agencies who cannot tolerate the costs or time requirements of additional upfront research.

WHY ADAPTIVE FIELD-TESTING?

The need for adaptive field-testing arises from the recognition that most projects cannot or should not be designed as "blueprints" involving highly specific and detailed planning and scheduling prior to the beginning of implementation. The "blueprint" approach assumes that technology and intervention techniques appropriate for a particular undertaking are known and can easily be applied to a target population. It assumes the existence or easy creation of institutions capable of implementing a project. It assumes that all critical information gaps can and must be filled prior to implementation and that each and every activity to be carried out during the project can be specified, costed and scheduled in advance.

A contrasting view emphasizes that a design is often completed with several "unknowns" left hanging, but that the existence of these "unknowns" should not delay the start of implementation. This approach emphasizes that predicting social dynamics is difficult at best, that design and implementation should be seen as a continuous process rather than two separate activities, and that it is frequently better to eliminate the "unknowns" within the context of project activities. Project development, in other words, continues into the implementation stage and involves a process of systematically eliminating "unknowns" through adaptive field-testing.

The benefits of adaptive field-testing accrue to both project management and project participants. For management it provides a way of acquiring information to help in decisionmaking and to acquire the information under management control. Many development projects require mid-project evaluations to determine which elements of the project to continue, discontinue or revise, according to the results. In evolutionary projects, third- and fourth-year budget obligations may depend on a favorable evaluation at the end of the second year. Adaptive field-testing conducted during the first and second years can provide project management with a sound and relevant data base as well as ideas for alternative interventions or technologies that will assist in the preparation of a constructive evaluation. Frequently this can be done at very low cost and without the enormous time lags and information slippages that often result when research is conducted independent of the project.

Adaptive field-testing may also be seen as a project activity with its own intrinsic development benefits. When done properly it involves a learning process that is of direct benefit to project participants. New facts and relationships about a local area are learned and old facts are cast in fresh molds. Participants also learn to look at the possibility of change in a more systematic way and to take part in formulating new activities. When project participants are involved in carrying out, even helping design, their own experiments, they will also be more likely to accept the validity and applicability of the results.

WHAT CAN BE TESTED?

Development projects consist of three elements: an intervention, an intervention technique or process, and a management structure. Each of these elements may be subject to adaptive field-testing, albeit in different ways.

An intervention is whatever the project is introducing to a sector, to an area, or to a target population. In a capital assistance project this may be a bridge, a dam, a road. In an education project it may be curriculum revision within a formal school system, literacy classes for adults, in-service training for civil servants. In an agriculture project it may be fertilizer, improved seeds, a new planting technique. Any one project may, of course, have several interventions; an IRD project may combine interventions from several sectors. Adaptive field-testing may help a project designer or manager decide which seed variety is most appropriate for his area, what interest rate to set for loans to small farmers, or what optimal planting times are for new agricultural practices.

An intervention technique or process is the way in which the intervention is introduced. It may involve selection of an area or population, selection of intervention agents, a method of organizing people, or choosing among communication techniques. Should a new planting technique, for example, be introduced through young extension agents or through paraprofessional

farmers? Should it be introduced to individual farmers, to key farmers, to farm families or to groups of farmers? Should demonstrations be given at a farm center or in villages? Experienced project personnel know that development projects run into trouble from improper intervention techniques as often as from inappropriate technologies. Adaptive field-testing can be used to clear up some of the uncertainty with respect to these and similar questions concerning the most effective intervention techniques.

A management structure is a way of organizing project participants, channeling money and scheduling activities so as to maximize the efficiency with which project activities take place. The organizational structure of a project is frequently complex, involving several host country agencies, sometimes more than one donor organization and a hierarchical authority and decision-making structure reaching from the national capital through regions and districts to the village level. Project designers often have little choice of management structure, especially when a particular organization has been preselected by the host country government to manage the project. In some cases, however, adaptive field-testing may be used to test the appropriateness or effectiveness of lower-level structures. An experimental approach may be used, for example to decide through which organizations to work or how operational responsibility can be divided among organizations or within the same organization.

CHAPTER TWO

SOCIAL EXPERIMENTATION AND ADAPTIVE FIELD-TESTING

Let us distinguish between two ways of looking at experimentation at the project level:

- The project, or some aspect of it, as an experiment in and of itself; or
- A dynamic ongoing process where the project serves as a time-and-resource framework for systematic testing <u>and</u> application of techniques identified as appropriate to local conditions.

The first view embodies the idea of social experimentation, i.e., systematic testing of interventions for the purpose of planning (or evaluating the effects of) various elements of a development project. The second view comes closer to capturing the idea of adaptive field-testing, since it emphasizes the notion that testing may take place during, and as an integral part of, project implementation.

There is a great deal of overlap between these two views of project-level experimentation. A discrete test conducted within a project may come close to following all the procedures of social experimentation, but at the same time that test may be seen as part of a larger integrated learning process involving the project as a whole. It is useful, nevertheless, for the project designer or manager to have some notion of the elements

of social experimentation. He will then be in a position to see what role experimental techniques may play in his project and how they may have to be adapted to field conditions.

In this chapter we will briefly review the elements of social experimentation, and then see why adjustments are needed for adaptive field-testing.

THE ELEMENTS OF SOCIAL EXPERIMENTATION

Experimentation may be seen as having three phases, each of which is composed of several steps or elements. The three phases are:

- Problem definition and formulation of hypotheses;
- The experiment; and
- Interpretation of the experimental results.

Problem Definition and Formulation of an Hypothesis

The problem is generally stated in terms of the need to reduce uncertainty about the consequence of some activity. Examples are:

1. Will fertilizer increase yields sufficiently to justify its cost?
2. Will a group credit scheme work in a particular society?
3. Can farmers' sons be recruited, trained and effectively employed as extension agents?

The next step is to state the problem in the form of an hypothesis. Hypotheses generally take the form of positive statements alleging a relationship between a particular factor and some desired effect. Experiments may be set up to test several hypotheses or, more frequently, a study may include several experiments, each of which tests a single hypothesis. Examples are:

1. The use of urea-based fertilizer will increase coffee yields sufficiently to justify the cost of the fertilizer.
2. A group credit scheme will attract more borrowers to invest in the new technology than will an individual credit scheme.
3. Farmers' sons, once trained, will be able to persuade their families to adopt improved agricultural practices.

The Experiment

An experiment is a contrived activity set up to test an hypothesis. It consists of one or more treatments administered to some set of persons or other units drawn at random from a specified population; and of observations or measurements made on the effects of the treatment to learn how (or how much) the treatment has caused the treated persons or units to be different from some untreated or control group that has been drawn at random from the same population.

The Treatment

The treatment is the subject of an hypothesis. It is the fertilizer whose effect on yields we want to test. It is the group credit scheme. It is the use of farmers' sons as change agents. (This factor being tested is sometimes referred to as the active treatment to distinguish it from an alternative or dummy treatment administered to a control group.)

Units of Observation or Measurement

The essence of an experiment is to "treat" some set of units (persons, hectares, etc.) and to compare the results of that treatment to the results of an alternative treatment (or no treatment) applied to an identical (or very similar) set of units, i.e., the control group. Thus the next task is to decide on the units that in one group will receive treatment and in the comparison or control group will not.

Certain conditions govern the choice of units of observation. First the units must be identical or quite similar to each other in all important characteristics. As an alternative, or in addition, there must be enough of them so that, in drawing a sample, the chance of assembling a treatment group and a control group that are dissimilar in some important characteristics is minimized. If, from a group of 20 villages, two have been chosen to compare the acceptance of a group versus an individual credit scheme (where one village would receive group credit and the other individual credit), then the two villages must be as

identical as possible in all characteristics that might affect the acceptance and use of credit. Such characteristics might include the relative wealth and educational level of the villages, current access to alternative sources of credit, and past experience with institutionalized credit. If the purpose of the experiment goes beyond merely comparing the relative acceptance of the two credit schemes to estimating the probable acceptance of the schemes in the other 18 villages, then the two pilot villages must also be "representative," or similar in all important characteristics to the population as a whole, which in this case is all 20 villages. If the 20 villages are so dissimilar in aggregate characteristics that choosing two that are representative is not possible, then it may become necessary to change the unit of measurement. Rather than measuring acceptance by village, we might measure acceptance by individual farmers.[1]

Random Sampling

The purpose of an experiment is to isolate and test the effect of a particular factor -- the (active) treatment -- on a particular population. In order to be sure that observable effects are due only to the influence of the treatment, it is important that all other factors -- potential influences -- be held constant. The most common way to accomplish this is to

---

[1] Actually, even when the unit of measurement is the village, we would start by observing the reactions among individual farmers. These observations would then be aggregated by village, since it is the collective reaction of the village rather than individual reactions of the farmers (in this case) that we are interested in.

use a sampling technique -- usually random sampling -- that will ensure that the variance in critical characteristics in the experimental group will be as similar as possible to the variance in the control group. By picking a large enough sample at random from the population and by dividing the units equally and at random between experimental and control groups, we can generally be confident that outside factors will have a neutral or constant effect on the experiment.[1]

Experimental and Control Groups

Once the sample is drawn the next step is to divide the sample into two groups: an experimental group and a control group. The experimental group is the group subjected to an active treatment, often referred to simply as the treatment. The control group receives a control treatment often amounting to no treatment at all. The purpose of the control units and control treatments is to permit the observer to measure the effects of the (active) treatment and to draw an inference or conclusion about whether (and possibly how much) the (active) treatment has affected the experimental group.

Observing the Effects of the Treatment

The next step in an experiment is to administer the treatment or treatments to the experimental group and the control group and to observe or measure the effects of the treatment.

---

[1] Randomization is neither always possible nor always desirable. Other methods of drawing a sample will be discussed in Chapters Three and Four.

Examples are:

1. Fertilizer is used on the experimental plots; no fertilizer is used on the control plots; yields are recorded in both groups.

2. Individual credit is offered in one village, group credit in another; the number of borrowers in each village is recorded.

3. Farmers' sons are used as extension agents in one group, the use of state extension agents is continued in another; the rate of adoption of improved agricultural techniques is observed in the two groups.

The period of time that will be allowed for the treatment to take effect will, of course, depend on the nature of the experiment, but in every case it should be stipulated in advance.

Recording and Analysis of Differences

The final step in the experiment is to determine the difference in the recorded observations or measurements. If the experiment has been properly conducted, the difference in results (if there is any) can be attributed solely to the influence of the treatment. Examples are:

1. If the fertilized fields produced 100 kg. per hectare more than the unfertilized fields, then, everything else being equal (and this is what the experiment should control), the fertilizer is responsible for the difference.

2. If 50 percent more farmers accepted credit when credit was introduced through a group lending scheme than through individual lending, then (everything else being equal) the use of the group credit scheme is responsible.

3. If more farmers adopted new techniques when the techniques were demonstrated by their sons

than when they were demonstrated by state extension agents, then (all else being equal) the use of farmers' sons as extension agents is responsible for the difference in adoption rates.

The magnitude of the observed difference is frequently important. In general the larger the difference, the less likely it is that it resulted from some factor other than the treatment. Statistical formulae are available for calculating the probability that a non-treatment factor caused the observed difference. It is conventional to assume that the results of an experiment are significant if there is less than one chance in 20 that the observed difference is not the result of the treatment.

Interpretation of the Experimental Results

If analysis of the experimental results reveals that an observed difference is statistically significant, that it is almost surely the result of the treatment, the next step is to decide whether it is practically significant. Examples are:

1. An experiment has demonstrated that the use of fertilizer has increased yields by 100 kg. per hectare. But will the farmer have to buy the fertilizer on credit, thus increasing costs while adding another element of risk to his enterprise? Will the farmer incur other costs connected with the use of fertilizer, such as transportation, storage and labor? Will the farmer have to change his planting time, adopt a new crop mix, change his cultivation techniques or accept a new marketing pattern?

   All things considered, will the yield increase of 100 kg. per hectare justify these direct and indirect costs?

2. An experiment has demonstrated that the availability of group credit for farmers has attracted a greater number of borrowers than when credit was available only on an individual basis. But what are the differences in administrative costs between group and individual lending? Will the default rate be higher, lower or the same? Which method of lending would be more likely to attract repeat applicants?

   How should these considerations be balanced with the proven initial attractiveness of the group lending scheme?

3. In an experimental situation farmers' sons have proven more effective than state extension agents in getting traditional farmers to adopt improved methods. But what are the constraints involved in widespread recruitment of farmers' sons? Once trained, can farmers' sons be effectively supervised? And how long can they be expected to remain in the villages?

   Will the practical answers to these questions mitigate against using farmers' sons as extensionists despite their proven effectiveness?

The results of adaptive field-testing can produce information that reduces certain elements of uncertainty that might be involved in a decision. But experimentation cannot substitute for the decision itself. The decision can only be made by human beings using interpretation, judgment and common sense.

THE NEED FOR ADAPTATION

The experienced development practitioner, having read the last section, will probably be thinking of a thousand reasons

why the process described in that section is not practical and cannot be used in the field. This thinking is realistic and valid. Most development projects operate in situations that reduce, if they do not actually eliminate, the possibility of conducting social experimentation as it is classically conceived.

The procedures of social experimentation and the process of project development begin at the same point. They both aim to clear up unknowns with respect to the environment and to reduce the level of uncertainty around which important decisions must be made. Critical differences arise, however, in the degree of control that development practitioners are actually able to exercise, at reasonable cost, over the environment they work in. The selection of field-testing sites, for example, comes about only through interaction between project developers and other interested parties -- including host country authorities, donor agency officials, and (at least in theory) members of the prospective target population. Hypotheses to be tested may be formulated implicitly rather than explicitly, and often after experimentation and data-gathering have begun. The attempt to ensure equivalence between experimental groups and control groups through randomization may prove difficult and costly, if not practically impossible. These real-world situations (and there are many others) force the prospective development experimenter to tailor his strategy to specific circumstances.

At the beginning of this chapter we posited two ways of looking at experimentation at the project level:

- The project, or some aspect of it, as an experiment in and of itself; or
- A dynamic ongoing process where the project serves as a time-and-resource framework for systematic testing and application of techniques identified as appropriate to local conditions.

The first view is more consistent with the procedures of classical social experimentation. But it implies a degree of rigor and precision that can rarely be achieved in the context of a development project. The second approach better captures the idea of adaptive research. It acknowledges the lack of control available to the experimenter and recognizes that even "quasi-experimental" or "second-best" procedures may not be feasible, either for lack of definitive hypotheses or due to an inability to identify and control for all the variables potentially influencing the problems under investigation.

The second approach does not concede, however, that "experimentation" under such conditions is not necessary or will not produce useful results. Given the urgency of the problems confronting target populations in many project situations, identification of techniques that work may justify full-scale application even if the reasons that they do are not fully understood. For example a great deal can be learned in a post hoc analysis of a "treated" population by investigating why some small farmers adopted a particular intervention and others did not.

The importance of finding and applying solutions within the framework of a development project determines another critical difference between classical social experimentation and adaptive field research. In adaptive field research, testing of a particular alternative is not pursued indefinitely, but only as long as it appears promising. For example, tests during a single cropping season of four to six months may yield enough evidence to justify dropping an agricultural input from the inventory of possible interventions. More complex tests (for example comparing the results of reliance on alternative methods of knowledge transfer between farmers) would require continuous monitoring over longer periods (perhaps even the full life of the project).

Depending on the focus of a project component, adaptive research may begin with a larger number of possible techniques, but the range will be progressively narrowed. So long as this process is purposive and systematic, it is logical to expect finally to identify optimal techniques for the project to introduce in the target area.

CHAPTER THREE

DEALING WITH UNCERTAINTY: SOME EXAMPLES FROM EXPERIENCE

We have argued that adaptive field-testing should be seen as a dynamic ongoing process for systematic testing and application of techniques identified as appropriate to local conditions. We have also conceded the necessity for frequent deviation from the rigorous procedures of classical experimentation. However, each deviation from classical procedures should be done deliberately, out of necessity rather than ignorance. In this way the costs -- e.g., potential loss of explanatory power -- as well as the benefits will be properly considered.

LEVELS OF UNCERTAINTY

Field-testing helps us make decisions under conditions of uncertainty. In deciding what type of field-testing might be necessary (or justifiable), the level of uncertainty is an important consideration.

"Blueprint" Designs

At one extreme a project designer or manager may be absolutely certain that the technology and the intervention techniques already identified are appropriate and, given goodwill

and skillful management, will work in the local environment. This point of view is consistent with the "blueprint" approach to project design and implementation. It assumes that solutions to the problems of development are known and that projects are merely vehicles for applying them. If projects are designed and managed using the blueprint approach, field-testing will be seen as unnecessary; hence none will take place. When blueprint projects fail blame is generally placed on "poor management" or "lack of cooperation" rather than poor selection of technology and intervention techniques.[1]

Formative Evaluations

At a middle point along the certainty-uncertainty continuum, we may be relatively certain about our technology and our intervention approach, but would like to test parts of it at an early stage of project development/implementation. We do not really have alternative inputs or approaches in mind; we simply want to be sure that the ones we expect to use will work. At this level of uncertainty we may want to engage in a variety of adaptive field-testing known as formative evaluation. An information sys-

---

[1] Recent research has shown that projects designed and managed under the blueprint approach run a high risk of failure precisely because they have not included information-gathering, field-testing and the flexibility to change directions on the basis of new information fed into the project itself. Yet the conventional blueprint approach remains quite popular with donor agencies and host country governments, who cannot be bothered with the time and expense of "more studies." See Development Alternatives, Inc., *The "New Directions" Mandate: Studies in Project Design, Approval and Implementation* (two volumes), prepared for the Agency for International Development, January 1978.

tem will be established and data will be collected to monitor the effects of the intervention or technology about which we harbor some uncertainty. As long as the feedback from the information system indicates that the intervention is coming reasonably close to having the intended (or some other beneficial) effect, we will continue with it. (The information system may also indicate that the overall intervention is sound, but that some minor tinkering will improve its impact.) If, on the other hand, the data indicate that the intervention is simply not working or is causing more problems than it is worth, we may be in a position to jettison it at an early stage in the project and to look around for alternatives.

It is clear that this type of formative evaluation differs markedly from post hoc or summative evaluations, which are more common in development projects. In a summative evaluation, conclusions about whether a given intervention has worked or not worked are frequently reached when the project is completed or when it is too late to change directions. With formative evaluations, information is gathered and observations are made during project implementation so that conclusions and decisions based upon them can be made in time to do some good.

Relative Uncertainty

Toward the other end of the certainty-uncertainty continuum, we may have little or no idea about what interventions or intervention techniques are most appropriate for our project. Two

reactions are commonplace at this level of uncertainty: (1) If the uncertainty revolves around technology -- the proper seeds, proper planting times, etc. -- the project may be changed to become a "research project." A research station will be established and trials conducted, often with little or no involvement on the part of the local population. At some later stage, once the "best technology" has been tested and "proven" in the research station, an attempt will be made to transfer it to a local population through another project, often of the blueprint variety. (2) If the uncertainty revolves around intervention techniques -- such as choice of change agent or methods of organizing farmers -- the opposite reaction is common. When project designers are unsure of how to proceed and wary of the reaction of donor agencies and host country governments, they often simply stipulate that a certain procedure will work and design it into the project. "Cooperatives are stipulated as the appropriate vehicles for organizing and mobilizing target groups in many agricultural and rural development projects -- not because we know that these arrangements will work, but because not much is known about workable patterns of collective behavior, and some pattern has to be established."[1]

It is possible to postpone the delivery of benefits while awaiting more research station results, or to ignore the problem

---

[1] PASITAM Newsletter Number 17, B[illegible]ngton, Indiana, Spring 1978, page 1.

of uncertainty by stipulating solutions. However, we argue here that, even at high levels of uncertainty, a development project that is intended to deliver benefits to a target population may often be an appropriate vehicle for resolving or reducing uncertainty. At high levels of uncertainty, adaptive field-testing in which alternative interventions are compared to each other is indicated. Rather than assessing the effects of only one intervention, as in formative evaluation, we would simultaneously compare the effects of alternative interventions. In making such comparisons, we would try to follow the procedure of classical experimentation as closely as possible, taking note of whatever deviations are necessary and assessing whether and how much these deviations might affect the conclusions we draw from the results of our testing.

There are obviously many points along the continuum, from absolute certainty to absolute uncertainty. Different elements of the same project, moreover, may fall at different points in the continuum. For elements about which uncertainty is low, we may content ourselves with some sort of formative evaluation procedure to monitor the impact of those elements. For elements about which uncertainty is high, we may want to conduct more elaborate field-testing. Various degrees and combinations are possible within any project.

The remainder of this chapter will give four examples of various types of adaptive field-testing. The examples will

illustrate ways of dealing with various levels and types of uncertainty under field conditions. In Chapter Four lessons derived from these and other examples will be systematically developed.

EXAMPLES OF ADAPTIVE FIELD-TESTING

Single-Factor Uncertainty

In the first example we will examine a situation in which we are uncertain about the use of one element in a technological package. The choice is between the entire package or the package without the element in question. Our treatment will be to administer all elements to the treatment group and to withhold the element about which there is doubt from the control group.

The Problem

A small farmer coffee project has been designed with a technological package that includes improved cultivation techniques -- elimination of old or diseased trees, reduction of the shade cover, no interplanting, etc. -- with the introduction and supervised application of chemical fertilizer. The problem is uncertainty about whether the fertilizer will increase yields enough to justify its costs.

The Hypothesis

Having identified the problem, the next step is the formulation of an hypothesis to be tested. In this case, the hypothesis would be that the fertilizer will indeed increase yields sufficiently to justify its cost. The hypothesis must specifically state how much of an increase in yields is expected and over what period of time. For example, in the first harvest after fertilizer is applied, fields that were treated with fertilizer will yield at least 200 kg. of dry coffee beans more per hectare than fields using no fertilizer. This hypothesis implies that any per hectare difference in yields inferior to 200 kg. would not be sufficient to justify the use of the fertilizer.

The Unit of Observation

The statement of hypothesis has also identified our unit of observation, or experimental unit: in this case, kg. of dry coffee beans per hectare. Notice that other units could have been chosen, e.g., kg. of dry coffee beans per farmer, kg. of dry coffee beans per farm, or even kg. of dry coffee beans per coffee tree. If one of the first two units had been chosen, however, two variables would have been introduced that would have been difficult to control through sampling procedures. The first is size of farm: since some farms have more hectares than others, differences in output per farm or per farmer in the experiment might be caused simply by greater or lesser hectarage rather than the use of fertilizer. The other variable that would

have been difficult to control is the skill of the farmer. If either farm or farmer were chosen as units of measurement, the difference in management abilities of the farmers in the sample might bias the pure observation of the effects of the fertilizer. The choice of kg. per hectare as the unit of measurement renders the size of farm largely irrelevant and farmers' skills can be more easily controlled for (either by choosing a large number of farms or by comparing the effects of the use and non-use of fertilizer on the same farm).[1]

By choosing hectares as the experimental unit, on the other hand, there is still one factor left to chance, and that is the number of coffee trees per hectare. It would not be valid to compare the yields of two hectares (one with fertilizer and one without) if one had more coffee trees than the other. Theoretically, then, it would be preferable to measure yields per tree rather than yields per hectare. However, counting trees and yields per tree is impractical; therefore we assume that, if there are enough hectares in the sample and if the hectares are distributed randomly between experimental group and control group, the number of trees in the two groups will be approximately equal.

There are a large number of such variables that we must assume will be randomly distributed between control and treatment

---

[1] We are assuming, for the sake of this example, that we are <u>only</u> interested in the effects of the fertilizer. In the real world other factors, including size of farm, might be of very real interest with respect to their influence on yields. Thus we might want to let size of farm be a variable rather than a constant in order to measure, rather than control for, its effect.

groups, e.g., size and age of tree, surrounding vegetation, access to sunlight, quality of soil, incidence of disease. In adaptive field-testing we should be aware of the possible influence of all these factors, even while recognizing that we cannot explicitly control for all of them.

Another factor implicit in the hypothesis is the way that the effect or outcome of the experiment is measured. Notice that the measure is not simply yields per hectare, nor even increase in yields per hectare, but the difference in yields (or the difference in yield changes) between two groups: one that used fertilizer and one that did not. In effect there are two comparisons that will be made here. One is internal to the experimental group: it will compare the yields before fertilizer to the yields after fertilizer on the same hectarage. The second comparison is between the experimental group and the control group: it will compare yields on the hectarage using fertilizer to yields on the hectarage where no fertilizer was used. Note that both groups would be subjected to the other improved cultivation techniques.

Sampling

Having identified the units of observation, the next step is to draw a sampling of these units and to divide the sample into an experimental group and a control group. There are several ways this can be done. The two most common are random selection and matching.

The preferred sampling technique, as explained in Chapter Two, is through randomization. In the present example two methods of random selection suggest themselves. One method would be to pick a certain number of farms at random from the overall population of farms, then randomly to assign half the farms to an experimental group and half to a control group.[1] Another method would be to select, again at random, a certain number of farms and to divide each farm into two equal plots: an experimental plot and a control plot.[2] Practically, the second method raises the problem of convincing each farmer to apply his allotment of fertilizer only to his experimental plot and then making sure he does so. However the second method is theroretically preferable because it controls effectively for two variables -- size of farm and the management skill of individual farmers -- that might, under the first method, rival the fertilizer in producing differences between the experimental and the control groups. Controlling for these two influences may also permit us to use a smaller sample.

The practical problem with both these random sampling methods is that they require a sample size sufficient to ensure that the distribution in the values of important characteristics in the experimental group will be similar to the distribution in the control group.[3] Under field conditions it may not be

---

[1] This would normally be done using a table of random numbers.

[2] On each farm the hectarage assigned to the experimental group and the hectarage assigned to the control group should also be selected at random.

[3] The problem of sample size is discussed in more detail in Chapter Four.

possible to work with a sample large enough to ensure such equivalence between groups. If we are limited in the size of our sample, an alternative sampling technique is matching. As the term implies, matching is a technique of deliberately (rather than randomly) assigning units of observation to experimental and control groups on the basis of one-to-one identity of characteristics.[1]

The Experiment

Once the sample is drawn we proceed to the administration of the experiment. Let us say that we have decided to take a random selection of farms and that our sample size is 100. Pretest, or baseline, data on the previous year's coffee harvest would be collected on all 100 farmers. Then, through random selection, the farms would be divided in half: 50 farms would become the experimental group and 50 farms the control group. In the experimental group the farmers would receive instruction in the improved cultivation practices <u>and</u> they would receive fertilizer to be applied to their fields under the supervision of extension agents. In the control group the farmers would only receive the instruction in improved cultivation practices.

At the end of a certain time period, e.g., one growing season, we would compare the yields of the two groups of farms. Each farmer, under the guidance of an extension worker, would record the yields on his farm and then the yields for each group

---

[1] See Chapter Four for further details on matching.

would be summed. The total for each group would then be divided by 50 to get the average output per farm. This figure would then be multiplied by the average number of hectares under coffee per farm to get the output per hectare. (We are assuming that average farm size is the same in both groups.)

Analysis

Let us assume that the output of the experimental group is 400 kg. of dry coffee beans per hectare (up 200 from the previous year) and that in the control group it is 300 (up 100 from the previous year). The next step is to analyze this observed difference of 100 kg. per hectare to determine whether it is statistically significant. Chapter Two explained that statistical significance refers to the likelihood that the observed difference could have been caused by any factor other than the experimental factor, i.e., the fertilizer. Several standard and relatively simple statistical tests are available to determine the significance. Some are simple enough that, if the numbers involved are not very large, they can be done by hand. Most can be done with the use of a hand calculator; very few require a computer.

It is generally not important for the project designer or manager to master these tests, since the statistical expertise needed to perform the calculations can be hired. However it is important for the designer or manager to understand, as common sense would warn, that observed differences are not necessarily caused by the experimental factor, that other factors may have

played a role. Statistical significance is a function of the magnitude of the difference and our tolerance for error. A small difference may be calculated as statistically significant if we are willing to live with the relatively high chance, say one in five, that some factor other than the experimental factor has intervened or caused the observed difference. If our tolerance is lower, for instance if we are only willing to put up with a one in twenty chance of random factor intervention, then the magnitude of the difference must be larger.

Our experiment thus ends with a determination of whether the difference of 100 kg. of dry coffee beans per hectare is statistically significant. If it is not, and if we are convinced that the experiment was properly conducted, we may conclude that the fertilizer has not significantly affected coffee yields. If the difference is calculated to be statistically significant, we may conclude that the fertilizer has caused greater outputs.

Interpretation

What comes next is a judgment about whether the increased output caused by the use of fertilizer is sufficient to justify including it in the technological package. This judgment must be made by using tools other than experimentation. These other tools may include cost-benefit and risk-benefit exercises, and analysis of the external market effects of fertilizer use or its effect on the environment. Experimentation has provided us with a valuable piece of information that we will then weigh

with the rest of what we know to judge the overall worth of the fertilizer in our project.

Regardless of the judgment we make with respect to the use of fertilizer, let us note that, by conducting the experiment within the context of an ongoing project, we have not postponed the delivery of benefits to a target population. Farmers in both the experimental and the control groups increased their yields. Although the non-sample farmers in the population received no benefits during the first growing season, it probably would not have been possible to reach all of them right away; and this majority of farmers, when finally reached, will benefit from the refinement in the technological package that was made possible through experimentation.

Controlling for a Single Environmental Factor

In some cases uncertainty revolves around the selection of target population, rather than the intervention itself. Using experimentation in these cases, the intervention becomes a constant while it is the selection of target population that becomes the treatment.

An example is provided by an actual design exercise carried out in Upper Volta.[1] The design team was uncertain about the effects of certain technologies for different village groups. The team believed that the feasibility of various interventions

[1] "Women's Roles in Development," Upper Volta, AID Project No. 686-0211, 1977.

would depend on the "level of development" of the villages. Level of development was defined in terms of two variables: the presence or absence of year-round access to water in each village, and the existence of dependable cash crops. This breakdown generated a classification of three levels of village development: Level One referred to those villages lacking both year-round water and a cash crop; Level Two designated those with water but without a cash crop, and Level Three encompassed those villages possessing both characteristics.

This three-tiered classification provided the framework for comparing the effectiveness of project activities (interventions) across different project environments. During the early stages of project implementation, the same activity would be introduced into the three different environments. The activity would be held constant so that the impact of level of development on the feasibility of the technology could be assessed.

Notice that such a framework also allows for pilot testing of activities at a particular level of development. We may be relatively certain that a technology will work in Level Three villages, and also quite certain that Level One villages are not "ready" for the technology, but we may be uncertain about the feasibility of the intervention in Level Two villages. Thus we can divide Level Two villages into two groups, an experimental group and a control group, to pilot test the activity. In this manner we are testing the activity itself while "controlling for" level of development.

Controlling for Several Environmental Factors

In the Small Farmer Coffee Project example, we spoke of the need to identify the factors that might influence the outcome of an experiment and to deal in some way with the variance in those factors among the population. We offered three ways of dealing with these sampling problems, two involving random sampling and another involving matching. The common denominator in all random sampling techniques is the assumption that the technique will result in an equal distribution of important population characteristics between experimental and control groups.

If we have reason to believe that random sampling will not result in equal distribution of characteristics between comparison groups, then we must "control" for the variance among these characteristics in the population in some other way. In the Upper Volta example the "level of development" variable was controlled for by simply dividing the population into three groups characterized by their level of development. Any comparison within one of these groups would then not be affected by the variance in level of development found in the overall population.

The Upper Volta case, however, dealt with only one variable. In many development situations there are several important variables that must be controlled for. As populations are carved up by more than one variable, moreover, possible combinations increase exponentially. As the number of possible combinations

increases, the number of units that fall into each group among a fixed population decreases. In order to get enough units into each subgroup, it becomes necessary to increase the sample size, sometimes to unmanageable proportions.

One way to deal with this problem is to use some combination of matched pair sampling with stratification, and then use rapid survey techniques[1] to reduce the number of subgroups that have to be dealt with. Suppose that a project environment contains four critical variables distinguishing members of the target population of small farmers, and that there are two values for each variable:

Variables	Values
Ecological zone	E_1, E_2
Farm size	S_1, S_2
Technology level	T_1, T_2
Crop mix	C_1, C_2

With two values for each variable, there are $2^4 = 16$ possible combinations.

Carving the target population into 16 matched groups, however, might be too costly and time-consuming. It could also generate an unwieldy sample size.[2] We need, therefore, to reduce

---

[1] This is discussed in greater detail in Development Alternatives, Inc., *Information for Decisionmaking in Rural Development*, submitted to the Agency for International Development, May 1978, Volume Two, Chapter Six.

[2] As a general rule sample size in each subgroup in such cases should be at least 30. Thus, 30 X 16 = 480, a sample size that might well be unmanageable under project conditions.

the number of groups. To do so we must know something about how the 16 combinations are distributed in the population.

It is possible, with the use of rapid survey techniques, to establish the general distribution of these variants. Let us suppose that a survey is carried out and it is found that five combinations account for ± 90 percent of the total. This would provide a basis for constructing a set of only five matched groups. It might also be found, as a result of a rapid survey, that the amount of variance within a particular feature being investigated is so low that it justifies dropping that variable from the set. Rapid surveys may also help to isolate the most disadvantaged group or groups of farmers, the $E_2S_2T_1C_2$ variant, for example, having the "problem areas" of continued dependence on an unproductive technology (T_1) and a crop mix (C_2) that offers little potential for increased income.

This procedure, in its broad outlines, can be applied in any project environment. The critical point of departure is agreement on the minimum set of variables, and from there the minimum set of values attached to them.

Initial Surveys and Multi-Factor Uncertainty

It has become increasingly common for complex development projects to be designed with an initial stage -- often referred to as a "pre-implementation" period -- specifically designated for information-gathering and testing of alternative interven-

tions and intervention techniques. These periods permit project managers to secure an information base and to experiment with different approaches to the development problem at hand without committing themselves upfront to any particular approach. They also allow project funders to get project activities underway at an early stage without committing themselves to long-term funding.

An actual case of such a project design is the District Planning and Rural Development Project (DIPRUD) in Atebubu District, Ghana.[1] Atebubu District is an area where there has been little development work and about which detailed information has not been systematically assembled. Thus the project designers felt that an initial period of one year should be designated for pre-implementation information-gathering and testing of intervention alternatives.

A major area for information-gathering and testing was in the agricultural sector. Based on the agricultural calendar of the District, shown in Figure 1, a period of three to four months was set aside for initial observations and surveys. These surveys would serve two purposes. First they would provide a basis for designing field trials of possible innovations on small farmers' own land, utilizing risk-sharing agreements between farmers and project management. Second they would delineate the optimal structure and content of a farm records subsystem within

---

[1] District Planning/Rural Development (Phase I), Ghana, AID Project No. 641-0073, 1977.

FIGURE 1

SIMPLIFIED AGRICULTURAL CALENDAR, PRE-IMPLEMENTATION PHASE:

ATEBUBU DISTRICT DEVELOPMENT PROJECT (GHANA)

Rainfall Cycle	Major Dry Season	Major Rains			Minor Dry Season		Minor Rains			Major Dry Season		
Month	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN
Yams, Year 1							•Planting →	→	→	→	→	
Yams, 2					•Harvesting →	→	→	→	→	→	→	→
Yams, 3	Harvesting Continues →											
Maize	•Planting 1st Crop →	→	→		•Harvesting 1st Crop →	→	→					
Maize							•Plant 2nd Crop →	→	→		•Harvest 2nd Crop →	→
Rice							•Plant →			•Harvest →	→	
Manioc	Planting and Harvesting Year-Round ←	←	←	←	←	←	←	←	←	←	←	←
Groundnuts			•Plant 1st Crop →	→	→	•Harvest 1st Crop →	→	→				
Groundnuts								•Plant 2nd Crop →	→	•Harvest 2nd Crop →	→	
Agriculturalist's Main Duties	Assemble survey team and carry out first rapid survey of agricultural environment.		Detailed survey of cropping patterns and practices in major agro-ecological zones. Identification of major constraints and risk perceptions.				Prepare Project Paper; with assistance of Information Systems Specialist, design farm records system for project.		Detailed preparation for field trials to start during Phase II (1979 planting season): set up village centers, coordinate procurement of materials, formalize risk-sharing scheme for participating small farmers.			

• Activity starting point ▬ Continual activity

the overall information system created by the project.

For purposes of illustration, let us assume that the surveys reveal that a priority agricultural problem in the district is that of yam decay and that post-harvest losses in this crop have been identified as a serious constraint, limiting the ability of farmers to market the crop successfully.[1] Let us also assume that the initial surveys, combined with information from scientific journals, have demonstrated that the decay is probably caused by some combination of inferior yam species and improper storage techniques. On the basis of this information we believe that an improved yam cultivar (species), combined with a better storage technique, could cut yam loss due to decay substantially.

This information generates two hypotheses:

- That cultivar "A" (not widely grown in the district) is substantially more resistant to decay than cultivar "B," which was found to be the most commonly grown variety of yam; or
- That a low-cost storage technique "X" (not currently in use on most Atebubu District farms) offers greater protection than "Y," which is the most commonly practiced method for storing yams.

A simple "quasi-experiment" can then be designed to determine the impact of interventions "A" and "X," whether introduced separately or in combination. It requires the designation of four groups of farmers:

---

[1] Yams are the major food crop of Atebubu District. At this writing, the DIPRUD project has not yet progressed into the testing stage.

- The A-X group will grow the new variety of yam and store it using new recommended techniques;
- The A-Y group will grow the new variety of yam and store it using traditional techniques;
- The B-X group will grow the traditional variety of yam and store it using recommended new techniques; and
- The B-Y group will serve as a control, growing the traditional variety and using traditional storage methods.

In order to guard against the possible influence of outside factors, the farmers in each group must be matched, i.e., they must be as identical as possible with respect to all characteristics that might affect the outcome of yam production and storage on their farms. The initial surveys would help us to identify the characteristics against which the farmers must be matched. These might include farm size (or area under yam cultivation), ecological zone, availability of manual labor (household and for hire), and possibly management skills.

If the variance in one or more of the particular characteristics is high, it is quite possible that random sampling might not give us matched groups unless we increase sample size to unreasonable proportions. If we must limit our sample size for practical purposes, we may have to abandon random sampling and set about deliberately matching the farmers in each group according to the characteristics we need to control for. Here our initial survey comes to the rescue again: it can quickly tell us whether our groups are matched and what corrections might have

to be made (substitutions, additions, etc.) to match them.

Once our groups are matched we proceed with the experiment. This consists of introducing the new variety and storage technique to the A-X group, introducing only the new variety to the A-Y group, introducing only the new storage technique to the B-X group, and introducing neither to the B-Y group. Each group then goes through a yam growing season (the same growing season controls for rainfall and most other natural environmental factors), yields are measured and compared at the end of the growing season, and yam loss from decay is measured and compared at the end of the designated storage period. Statistical analysis is then performed to see whether any differences among and between the groups are statistically significant.

This example is a simple illustration of how more than one factor can be tested in the same field experiment. Unlike the small farmer coffee example, in which only one intervention (fertilizer) was being tested, here two interventions (a new yam cultivar and a new storage technique) are being tested and being tested for their "interlinkages," i.e., their combined and simultaneous effects. Simultaneous testing of more than one factor increases the number of test groups that are necessary and may, in the real world of adaptive field-testing, limit the use of random sampling. Matched pair sampling is one of a number of alternative sampling techniques that can be used to control for the unwanted effects of outside factors.

The DIPRUD project was designed so that field tests would also identify problems that might arise in setting up a farm records system for the project. In conducting the yam storage test, a certain amount of systematic data-gathering would become necessary. Tests conducted on other crops, as well as on other facets of agricultural life in the District, would all constitute trial-and-error type experiments in the potentials and pitfalls of establishing and maintaining a systematic farm records system. The relationship between adaptive field-testing and the farm records system would, moreover, be symbiotic. As the farm records system becomes functional, it would identify more problems and, possibly, suggest solutions to those problems. These solutions could then become the subjects of further adaptive field-testing.

The DIPRUD project design team also recognized the potentially tenuous nature of the project's organizational arrangements. The purpose of the project is "to develop the ability of the Atebubu District Council and its supporting system of local, regional and national institutions, to involve the population in planning, management, implementation and evaluation of a self-sustaining integrated development program." The project was designed around this purpose in order to help strengthen the Ghanaian government's policy of decentralization. The project was, and is, seen as a pilot endeavor, testing the ability of local institutions to plan for and carry out their own development activities. Initial responsibility for project management

will be given to the District Council. However, precisely because local-level development has never before been systematically attempted, the Project Paper warned that it might become necessary to identify "alternative organizational arrangements for promoting local development." Details on how to maintain flexibility and identify alternative management structures will be worked out by the implementation team.

CHAPTER FOUR

METHODOLOGICAL OPTIONS

The examples in Chapter Three raised three methodological problems that continually arise in field experimentation. These are the problems of (1) research design, (2) sampling, and (3) data collection. In this chapter we will describe certain methodological options that may be available to project designers and managers with respect to these three problems and will discuss criteria for choosing among them.

RESEARCH DESIGN

Three Design Models

At the heart of any experiment is a comparison. We may compare the same group at two points in time, e.g., yields per hectare on a plot before the use of fertilizer and yields per hectare on the same plot after the use of fertilizer.

Model One:	Before	After
	O_1	$\circledast_1$

(In the following diagrams, "★" indicates a "treated" group.)

Or we may compare two different groups at the same point in time, e.g., two different plots, one having received fertilizer and one having received no fertilizer.

Model Two:	Before	After
Experimental Group:		$\circledast_1$
Control Group:		$\bigcirc_2$

Under ideal conditions we would seek to combine these techniques, e.g., we would measure yields per hectare on both the experimental group $\bigcirc_1$ and the control group $\bigcirc_2$ prior to the test and again measure yields on the "treated" experimental group $\circledast_1$ and the "untreated" control group $\bigcirc_2$ at the conclusion of the test.

Model Three:	Before	After
Experimental Group:	$\bigcirc_1$	$\circledast_1$
Control Group:	$\bigcirc_2$	$\bigcirc_2$

The power of this combined model of research design is that it controls for the possible influence of other factors over time.

The Uses of the Models

The third model is the preferred model, but frequently under field conditions it is either the first or the second model that we are forced to adopt. Yet both the first and second models are not unacceptable for most project situations. While the limitations of each should be kept in mind, project designers and managers should not hesitate to use one or the other when circumstances dictate. Moreover there are ways to enhance the explana-

tory power of each.

The second model, which compares two groups at the same point in time without benefit of any pretest or baseline data, is often encountered in midproject evaluations. After one growing season, for example, project management decides it is time to "take stock," to assess the effects of the fertilizer used in the first year. The problem is that, while good data exist on the yields after the first year, no one bothered to record yields in previous years and/or there is no aggregate agricultural census data that can be used to measure change over time. The data on hand permit a comparison between the yields on fertilized and unfertilized fields, but the lack of baseline data makes the comparison risky because we are not sure whether the two groups started from the same point, i.e., whether yields were comparable when neither used fertilizer.

Under these circumstances, there are still several ways to "make do." One is simply to accept that differences between the experimental groups and the control group are valid, judging perhaps that the differences are large enough that the fertilizer simply had to have made a difference. To the extent that the control group and the experimental group are similar, we may be relatively confident that the observed first-year differences are attributable to the treatment, even without pretest data. Another way is to take a small group of farmers from each group and ask them to recall their yields in the previous year (or perhaps their average yields in the five previous years). These retro-

spective data cannot be considered as reliable as if direct observations had been made of yields in previous years, but they are better than none at all. As a temporizing move, we might also accept the differences in the two groups as tentatively valid and use these observations as pretest data for corroborative tests to be conducted in the following year.

Model One designs, in which there is a comparison over time but without a control group, occur even more frequently under project conditions. For political or ethical reasons we may not be able to withhold treatment from one group while administering treatment to another: all the farmers in the target population <u>must</u> get fertilizer in the first year. Or we may be sufficiently certain that our intervention will work that we want to get it to the entire population as quickly as possible. Under these circumstances we can measure changes over time, but cannot, for lack of control groups, be sure that the change was brought about by the intervention.

Our best bet under these circumstances is to rely on a variation of Model One called time-series observations. Rather than making observations (measurements) at only two points in time (before and after), we make a series of observations at successive points in time:

Time-Series Observations:

Time 1	Time 2	Time 3	Time 4....
O_1	$\circledast_1$	$\circledast_1$	$\circledast_1$....

Using this model we can compare the observations at different points in time, e.g., at the end of each of four or more growing seasons. The difference in yields between Time 1 and Time 2 can then be compared to the difference in yields between Time 1 and Time 3, and so forth, to double-check the validity of the first set of observations.

The use of time-series observations is a type of formative evaluation. Such observations are facilitated and rendered more reliable when made through the establishment of a farm records system within a project. The farm records systems may in turn uncover new problems or potentials that may become the subject of future time-series observations or other adaptive field-testing.

SAMPLING

There are three questions concerning sampling techniques that frequently arise under field conditions. These concern (1) representativeness, (2) sample size, and (3) adaptive sampling techniques.

Representativeness

A frequently asked question is: does my sample have to be representative of some larger population? The answer is that it depends on the purpose of the experiment. If the purpose of the

experiment is simply to measure the effect of a treatment, then the sample does not have to be representative. The only requirement is that the observation units in the experimental and control groups be similar.

When scientists experiment with rats in a laboratory, they may be primarily interested in the effect of the treatment on human beings, but they will settle for being able to measure the effect of the treatment per se rather than the effect of the treatment on the group of ultimate interest. By observing the effects of treatment on two groups of rats, they may suspect, but they cannot infer (let alone conclude) that the treatment would produce similar results for human beings.

In like manner it is not valid to infer that, because an experiment or program has had a positive outcome among one group of people, it will have a positive outcome on dissimilar population groups or environmental conditions. Yet this is precisely the assumption that is made in many development projects. A certain seed variety of maize triples yields in Mexico, so it is introduced in Zaire. Cooperatives are well received in Bolivia, so they are introduced in Haiti. Land grant college extension techniques have worked well in the United States, so they are spread around the world.

The only way to know for sure whether any of these interventions will work in a given society is to test them on a representative sample from the population that we want to benefit.

Sample Size

To ensure that a sample is representative of some larger population, it is important that the units in the sample possess the same characteristics, or the same variance in characteristics, as the overall population. To achieve representativeness in a sample, it is necessary to know what characteristics are important and to know something about the variance in those characteristics.

If you are a project designer or manager, it is probably not important that you master the statistical and technical formulae used by statisticians to draw up a sample. Under field conditions it is unlikely that these formulae can be used with any precision. If they can be used you will most likely be able to hire the needed expertise.

There is one principle, however, that it is important for project designers and managers to understand. It is that sample size bears little or no relation to population size. Rather sample size is a function of the variance in important population characteristics. To illustrate, if the statisticians tell us that 50 farms are sufficient to represent the variance in characteristics in a population of 1,000 farms in a certain area, then it is quite likely that these 50 farms would do quite nicely to represent 10,000 farms from the same area as long as the additional 9,000 farms have more or less the same characteristics as the original 1,000, and as long as the variance in those

characteristics is more or less the same. Failure to understand this has led to a great many instances of wasted time and effort as larger samples than necessary are selected.

Adaptive Sampling Techniques

Even when these principles are known and appreciated, it may not be possible to apply them. In the first place the distribution and variance in important characteristics may not be known. If they are known statisticians may calculate a sample size that is not possible to fill. These circumstances call for adaptive sampling techniques.

Under the field conditions likely to be found in a rural development project it may not be possible to get the information necessary to calculate the variance in important factors. There may have been no survey of farm size in the area and management skills of the individual farmers may be difficult to quantify even if they are known. Thus adaptive field-testing calls for the experienced designer or manager to use his knowledge of the area (or the knowledge of others that he may have to draw upon) to estimate the variance in these factors in the area.

An optimal procedure might be to select a sample (of let us say 100 farmers) that seems intuitively sufficient to represent the variation to be found in the total population. People with knowledge of the local area could then check the sample to see whether it seems to represent the entire population. If the

sample seems, for example, to have a smaller percentage of large farms than the overall population, then another 25 farms could be drawn at random to see whether the new total, 125, might correct the perceived imbalance. This procedure would be continued until the sample is seen by local experts to represent the total population adequately, or until we decide that, given the costs of continuing this procedure (and of dealing with the larger sample), we could tolerate whatever imbalance (statisticians call it skewness) remains.

But what if we cannot, for practical reasons, deal with a sample of 125? What if we cannot rely on random sampling to achieve the desired representativeness? A common sampling technique in these circumstances is the use of matched pairs. Using this technique we deliberately (rather than randomly) assign units of measurement to the experimental group and to the control group on the basis of their characteristics. A large farm goes in the experimental group; a large farm must go in the control group. A farm at 500 meters elevation goes in the experimental group, so a farm at 500 meters elevation must also be included in the control group. A farm whose yields were high last year goes in the experimental group; a similar farm must go in the control group. This process continues until some balance has been achieved between the constraints of sample size and the inclusion in both experimental and control groups of units of measurement that represent and are matched on all the important characteristics of the population.

DATA COLLECTION

The success of an adaptive field-testing program will depend, in large part, on the systematic, timely and efficient collection of field data. To be effective adaptive field-testing should be conceived as an integral part of a project information system and the results of individual tests should be seen as filling information gaps for project decisionmaking. The type of data to be collected will depend on decision categories that need to be addressed, information gaps that need to be filled and the nature of the field tests to be conducted. To prevent data overkill, it will be helpful to recall that data are only useful if they can be turned into information and information is only useful if it can be used in decisionmaking.[1]

Data Collection Strategies

Like other methodological decisions in adaptive field-testing, the choice of data collection strategy will depend on project circumstances. Three commonly used strategies are:

- Statistical surveys;
- Farm records; and
- Reconnaissance surveys.

---

[1] Guidelines for the selection of data points and information categories to be addressed in project design are presented in Development Alternatives, Inc., *Designing Projects for Rural Development*, submitted to the Office of Rural and Administrative Development, AID, August 15, 1978; see also, Development Alternatives, Inc., *Information for Decisionmaking in Rural Development* (two volumes), submitted to the Office of Rural and Administrative Development, AID, May 22, 1978.

Statistical Surveys

This term refers to techniques that utilize data from a sample to make inferences, i.e., to generalize about the characteristics of a larger population from which the sample has been drawn. Most commonly the statistical survey depends on enumerators who have been trained to administer a questionnaire or comparable collection instruments, with predetermined categories of data. Two principal types of statistical surveys used in rural development are the area frame sample and the population sample.

- The area frame sample utilizes a specified geographical area -- usually a small segment or "block" in the total land area of a region or country -- as the unit from which desired data are to be collected. Generalizations about agriculture and/or other economic activities within the total land area are derived by compiling data gathered within the selected segments.

- Population sampling is used when the focus of the survey is on a particular target population or on specific categories within the population inhabitating an area. Here the basis of the sample is not territorial units but, rather, a given number of reporting units (e.g., households, individuals, farms) selected from the total number of such units.

Farm Records

This approach is intended to gather data on farm operations on a continuous basis over an extended period of time. Entries are normally made at very short intervals -- sometimes on a daily basis -- and this feature increases the quality and quantity of

the data entering the system. This in turn tends to limit the size of the sample, whether an area frame or some form of population sampling is used to generate it. Although farm records have been used most often for purposes of research, they have great potential utility within ongoing projects in that they can monitor changes at the micro level that result from project interventions.

Reconnaissance Surveys

This approach is considerably less structured and formal than statistically oriented surveys. It depends on an open-ended process of questioning and observation, conducted by one or more qualified rural development specialists who concentrate the collection effort on key informants (as opposed to "representative" respondents). The rationale underlying the reconnaissance survey assumes that it provides a way of synthesizing data rapidly into information, drawing on the analytical skill of the rural development specialists.

Any or all of these approaches could be used in a single rural development project. Chapter Three gave examples of how rapid reconnaissance surveys can be used to identify the distribution and variance in important characteristics for the purpose of drawing a sample. Such surveys might identify salient information that should be gathered in a baseline survey, or that might be included in a monitoring effort or a formative evaluation. Information identified in a rapid survey might also be used to help structure a farm records system.

Data Collection Techniques and Special Problems

Project managers will need to make three procedural decisions with respect to their data collection systems.

The first involves the resources needed for the identification of data points and the design of data collection instruments. It is here that local expertise, combined with some expertise in methodology, is needed. Social scientists -- expatriates or host country nationals -- who have worked in and are familiar with the local area are frequently an invaluable resource for this task.

A second decision concerns the choice of data collectors. If structured questionnaires are used in statistical surveys, locally recruited primary or secondary school leavers can often be trained to administer the questionnaires. The open-ended techniques involved in reconnaissance surveys, on the other hand, require experienced and highly skilled field personnel. The monitoring involved in a farm records system requires trained extensionists who are an integral part of the project staff. To the extent possible, data collection for adaptive field-testing and other information system needs should be done by project staff. For purposes of objectivity, on the other hand, the analysis and interpretation of data collected for evaluation purposes should be undertaken by a team that includes non-project staff.

The project manager must also decide on the extent to which he will utilize short-term adaptive field-testing specialists. The project will generally need short-term assistance early on, when initial strategies are being developed and data collection instruments designed. Any type of structured questionnaire should be pretested; the supervision of the pretest, and especially the analysis of its results, may require short-term expertise. As the need for additional tests are identified, as sampling or statistical analysis needs arise, or as special operational problems crop up, additional or continuous technical assistance may be necessary.

CHAPTER FIVE

INTEGRATING OPTIONS AND POSSIBILITIES INTO A RURAL DEVELOPMENT PROJECT

TIME-PHASING OF EXPERIMENTS

Complex rural development projects commonly present a complex set of "unknowns," each calling for various kinds of experimentation. Some of these may involve interventions, others the process by which the intervention is introduced, and still others the management structure of the project. Beyond methodological problems a project designer or manager may be at a loss to know how to start the experimentation program, how to sequence the experiments needed and when to stop.

It should be appreciated that "solutions" to complicated development problems are rarely crystal clear and unequivocal. Even well-conducted laboratory-style experiments usually end up with "probabilistic" answers rather than definitive solutions. Thus in dealing with problems of development and social change, we must frequently content ourselves with tentative results and possible solutions. This means that the process of experimentation -- in the larger sense of probing for still better answers -- is a continual and reiterative process. Experimentation leads to action, which leads to further experimentation. If we under-

stand that this state of things is the norm, we will not wait too long for just the right answer before proceeding with our program.

Experience with rural development projects over the last five to ten years has, on the other hand, revealed that there may be a certain logical process to follow in project experimentation, just as there is a logical process for the implementation of rural development projects. If we make a minimum assumption that the project aims to introduce some new technology in agriculture (the intervention), and has the overall objective of increasing agricultural production and productivity, the following paradigm for adaptive field-testing may prove useful:

Step One

> Determine the best locally available technology currently used by target area farmers and conduct adaptive field tests on:
>
> (a) That technology applied to farmers who use something less than the best available; and
>
> (b) The process of transferring the knowledge about the "best" local technology.

In many rural development projects, rather than attempting to introduce new (outside) technology immediately, it may be better to experiment with locally available technology. Surveys conducted during project designs have frequently revealed that there are great differences in the levels and types of technology being employed in local areas. In Shaba Province, Zaire, for

example, it was found that a local population had a number of different ways to grow maize. In Gamo Gofa Province, Ethiopia, some farmers were found to be using simple wooden sticks for hoeing while their neighbors across the river had metal tips on their hoes.

In such cases it is frequently possible to set up experiments that aim to introduce the best local technology to those who are not using it, while at the same time experimenting with ways to transfer knowledge about that technology. This would give a composite "experiment" with two sets of unknowns, similar to the discussion in Chapter Three. It would also set the stage for further learning about farmer production techniques, constraints and potentials, as well as about the most effective methods of delivering new ideas to the local target group. Since the technology to be delivered is already available in the local area, no great leaps of faith are necessary to convince farmers that such technology can help them.

Step Two

> Experiment with technology, crops or cropping systems, storage, or other techniques that are new to the local area.

This should be tried initially on a test area or plot, and then extended to farmers' fields in a single- or component-factor adaptive field test. Such an experiment will help tailor the recommendations to take account of the variations in the environ-

ment (climate, soil, rainfall, disease, sunlight) in the project area.

This testing can be initiated as the project first gets under way, since it will not deliver outputs for use in the project until the second or third cropping cycle (or later, depending upon how "basic" the research must be). However since Step One is ongoing, there is a "deliverable" being provided to the target population, even in the early stages before the high-yielding innovations have been tested and selected for use.

Step Three

> Extend the new technology to the target population, using what has been learned about knowledge transfer methods during Step One.

In extending the new technology, a number of constraints must be remembered, including the novelty of unfamiliar crops, differences in cultivation techniques, water requirements, and disease control. If Step One has been successful, there will be communications agents who have established credibility with the farmers for the larger, potentially more valuable innovations that follow. This step must be fully integrated into or supersede the adaptive field-testing of Step One. Once all of the farmers have obtained high output and income from the existing technology, renewed progress will only be possible through the introduction of new higher-order technology. That is the purpose of Step Three.

Step Four

> Utilize the increased demands for change called for in Step Three as the springboard for testing local organizations and marketing associations.

The project may need inputs to be purchased at wholesale prices -- a good reason for a farmers' input cooperative. The increased output of the farmers may generate a surplus that requires shipment outside the local area, and thus new marketing and transport arrangements. Using these newly identified requirements, adaptive field-testing can be conducted on the optimum arrangements for target population cooperation and group action.

This step, which is one of the most important as well as one of the most difficult, can be attempted earlier in the project, perhaps at the start. If so, the form and structure of the local organizations are likely to need changing, as the requirements for interaction with the outside world (inputs, technical assistance, marketing) increase.

ORGANIZATION OF RURAL DEVELOPMENT PROJECTS USING ADAPTIVE FIELD-TESTING

Rural development projects that are designed to make maximum use of adaptive field-testing are often divided into components. Depending upon the needs of the project, the following grouping of activities has proved useful in the past:

- Research;
- Extension;
- Local organizations;
- Administration and finance;
- Infrastructure;
- Marketing and credit; and
- Data collection and analysis.

The Data Collection and Analysis Unit provides the overall guidance for the formative evaluation aspects of the adaptive field-testing, and assists all other project subcomponents in undertaking the testing necessary to improve knowledge and performance of the project as each cycle unfolds. There are obviously combinations such as grouping research with extension (which would be preferable in a small project), to reduce the number of Assistant Project Directors. Designed in this manner, adaptive field-testing can be made an understandable part of each subcomponent, and will not totally consume the time and attention of the project manager.

CHAPTER SIX

TEN QUESTIONS

INTRODUCTION

The first five chapters have described various processes and problems that will be encountered in adaptive field-testing. This chapter summarizes the salient features of adaptive field-testing in question and answer form. It can be used by project managers as an outline checklist in constructing their own adaptive field-testing program.

Concrete answers to the specific questions posed in this chapter can only be formulated in the context of a particular project. However, each question is followed by generalized options and criteria for choosing among those options. After each question, readers are also referred back to various parts of the first five chapters.

QUESTIONS

1. Should Adaptive Field-Testing be Part of My Development Project?

The answer to this question is, obviously, either yes or no.

The answer will be no if you are certain that all components of the project as designed are appropriate to local circumstances and will have the effects intended as stated in the objectives of the project.

The answer will be yes if you harbor uncertainty about the effects of one or more of the specific components of the project.

See Chapter One, pages 4-5 and Chapter Three, pages 21-26.

2. Which Elements of the Project Should I Experiment With?

Here, the answer will clearly depend on the nature of the project and its particular components as well as the level of uncertainty that obtains with respect to each component. However, it helps to be reminded that projects can generally be broken out into three elements:

- The interventions;
- The intervention techniques or processes; and
- The management structures including local organizations.

See Chapter One, pages 6-7.

3. How Rigorous Do I Need to Make My Field-Testing?

The choices here range from monitoring the effects of project activities (formative evaluation) through strict adherence to the classical precepts of social experimentation, with many points in between.

The choice depends on the level of uncertainty combined with the alternatives at hand. If the level of uncertainty is not high, formative evaluation is indicated. If the level of uncertainty is higher and/or if it is necessary to choose among several options, more rigorous experimental designs are preferred.

See Chapter Three, pages 21-26.

4. What Type of Experimental Design Should I Use?

The main choices are:

- Before and after;
- After only with control group;
- Before and after with control group; or
- Time-series.[1]

The choice will depend on:

- The availability of pretest data;
- The availability of comparable control groups;
- Political/ethical considerations concerning the deliberate withholding of a treatment from a control group; and
- Resource availability (Are the resources available, for example, to get data from a distant control group as well as from the target population?).

See Chapter Four, pages 47-51.

---

[1] There exist more complex designs but they are generally not suitable for adaptive field-testing.

5. Do I Need a Sample?

No, if the testing technique involves formative evaluation or if the target population is so small that all members can be used in the experimental group.

Yes, if the target population is large and/or if pilot testing on a small group seems desirable before extending an intervention to a larger group.

See Chapter Two, pages 13-14 and Chapter Four, page 57.

6. How Should I Choose My Sample?

The type of project will determine a choice between:

- Area frame sampling; or
- Population sampling.

Methodologically, the most common sampling techniques are:

- Random sampling;
- Random sampling with stratification;
- Matching; or
- Matching with stratification.

The choice of technique depends on:

- The number of units available for sample selection;
- The need to control for intervening variables; and

- The availability of local knowledge concerning the distribution and variance in important population characteristics.

See Chapter Two, pages 13-14, Chapter Three, pages 29-31, 34-78 and 42-43, and Chapter Four, pages 51-55 and 57.

7. Does My Sample Have to be Representative?

No, if it is only the effect of the treatment that we are concerned with.

Yes, if we want to know the effect of the treatment on some population larger than that contained in the sample.

See Chapter Four, pages 51-52.

8. If There is Uncertainty About Several Elements of My Project Where Should I Start?

A general paradigm for the time-phasing of adaptive field-testing was presented in Chapter Five.

9. How do I Know When to Stop?

In general, the answer to this question lies in the trade-off between the costs (in terms of time and resources) of further testing and the level of certainty already obtained.

If several technological options are under consideration, testing of each continues only so long as it seems to be having a positive effect.

See Chapter Two, page 20 and Chapter Five, passim.

10. What Practical Considerations Are There in Adaptive Field-Testing?

- The availability of capable (or trainable) data collectors;
- The money available to pay these personnel;
- The attitudes of host country officials toward a field-testing program;
- The pressure for widespread and immediate "results"; and
- Ethical considerations.

See Chapter Two, pages 17-19, Chapter Four, pages 59-60, and Chapter Five, passim.