

PW-AAQ-057

34707

PRINCIPLES AND GUIDELINES FOR THE EVALUATION OF
P.L. 480 TITLE II (FOOD-FOR-PEACE)
PROGRAMS AND POLICIES

Submitted to

Agency for International Development

Under Contract Number

AID/SOD/PDC 0262-I-00-1010-04

by

David E. Sahn

Community Systems Foundation

Ann Arbor, Michigan

October 1, 1981

CONTENTS

ACKNOWLEDGEMENTS v

Chapter

I. INTRODUCTION 1

II. CONCLUSIONS AND RECOMMENDATIONS 4

 A. A Proposed Framework for Future Evaluation Efforts 4

 B. Other General Recommendations 9

 C. Nutritional Impact 11

 D. Food-for-Work 13

 E. Educational Objectives 15

 F. Economic Impacts 17

 G. Integrated Programs and Related Impacts 19

III. PROCESS EVALUATION 20

 A. Project Context 21

 B. Program Objectives and Plans 22

 C. Service Delivery 24

IV. NUTRITIONAL IMPACT 27

 A. Previous Evaluative Efforts 27

 B. Alternative Approaches to Measuring Nutritional
 Impact 48

V. FOOD-FOR-WORK PROJECTS 54

 A. Process Evaluation 55

 B. Impact Evaluation 59

 C. A Related Note 64

 D. Nutritional Impacts 67

VI. ECONOMIC IMPACTS 71

 A. Microeconomic Effects 71

 B. National Economic Aggregates 77

 C. Inflation 78

 D. Production Disincentives 79

E. Economic Development	81
VII. EDUCATIONAL IMPACT	84
VIII. NUTRITION EDUCATION AND OTHER IMPACTS	87
.
APPENDIX A	91
APPENDIX B	93
REFERENCES	98

ACKNOWLEDGEMENTS

I wish to express special appreciation to my colleague, Roy I. Miller, for all his time in discussing and reacting to the ideas and concepts presented in this report. His assistance in critically appraising the analytic techniques used in evaluating nutritional impact is especially noteworthy.

I. INTRODUCTION

In 1977 the Congress of the United States passed a law which mandated the comparative cross-country evaluation of P.L. 480 (Food for Peace) Title II programs, to be performed in five-year intervals beginning on October 1, 1978. Emphasis was to be placed on assessing "the nutritional and other impacts, achievements, problems and future prospects for programs."

It can be surmised that the impetus for the Congressional demand to evaluate was predicated on the need to justify fiscal expenditures in an era of budget restraints and diminishing surplus food commodities. Providing evidence that the Title II program is effective in meeting its goals and in accounting for the costs involved in such an effort undoubtedly was the prime mover of the present enthusiasm for evaluation.

In the past few years there has also been a greater awareness among program funders and implementing agencies about the role of evaluation in providing information with which to make decisions (e.g., allocation of resources), as well as to assist in improving program performance. This trend is not unique to food aid and nutrition programs. But it was all the more compelling due to the inherent obstacles in documenting impact, and a long tradition of food aid being outside the purview of accountability (other than in terms of commodity flows).

As a result of these contributing factors, a number of P.L. 480 Title II evaluations have been performed since the Congressional mandate. Most have been initiated and funded by the Agency for International Development (AID). Some were performed under the auspices of the Food-for-Peace Office in Washington. The most recent of these was done in

accordance with a Generic Scope of Work, which is designed to guide the evaluation process. Other evaluations were a result of the urgings of other branches of AID/Washington (AID/W), such as the Office of Evaluation and the Regional Bureaus. In addition, some evaluations were performed due to the prerogative of USAID Missions in various countries, as well as the voluntary agencies themselves who design and implement Title II programs.

The thrust of this document is to delineate an approach to evaluation based on a review of the experiences to date. Emphasis is placed on issues surrounding impact evaluation. It is obvious that no clear and concise Agency-wide strategy for evaluating Title II programs has been developed. Different bureaucratic actors have different needs, and advocate different conceptual and methodological approaches to evaluation. At the same juncture, it is felt that many evaluations are asking the wrong questions. And when strategic questions about program impact are identified, methodological shortcomings have generally precluded the provision of unambiguous and convincing answers.

Therefore, the remainder of this report will critically review evaluations which have been performed previously (see Appendix A). In addition, principles and guidelines for the future of AID Title II evaluation activities are provided. It is stressed that in considering past conceptual and methodological difficulties, the general patterns and problems will be dealt with using specific illustrative experiences only when necessary. Scrutinizing and criticizing each evaluation individually was felt to be counter-productive.

In a similar vein, it must be recognized that this document is a first cut at re-orienting evaluation procedures. In some instances, the guidance provided will raise new questions, while answering others.

As such, the month-long consultancy which produced this report should be viewed as a beginning in putting the evaluation process in order.

The chapters that make up the remainder of this document are as follows: Chapter II is a brief summary of the recommendations and conclusions. By situating this chapter at the front of the text, it serves as a lengthy executive summary, in addition to orienting the reader to the material which follows. However, it does not substitute for reading the document in its entirety.

The next chapter concerns issues related to process evaluation. This chapter was included despite the fact that the interrogative focus of this report is the measurement of program impact. If nothing else, it will serve as a reminder that process evaluation is an integral and indispensable component of program evaluation.

Chapter IV is a discussion of nutritional impact as it relates to maternal and child health (MCH) and to a lesser extent school feeding (SF) programs. The evaluation of Food-for-Work (FFW) programs is the topic of Chapter V, followed by a discussion of economic impacts in Chapter VI. Chapter VII addresses educational impacts, and the final chapter focuses on conceptual issues related to measuring impacts (e.g., nutrition education) of integrated multi-service programs.

A final note to this introduction is that this paper assumes a level of knowledge concerning evaluation terminology and technique. Likewise, it assumes that the reader is familiar with the Food-for-Peace Program and/or similar large-scale food-aid activities such as the World Food Programme of the United Nations. And last, this document was prepared for AID personnel in Washington. Working knowledge of conceptual tools such as the Agency's Logical Framework will facilitate the understanding of this report.

11

II. CONCLUSIONS AND RECOMMENDATIONS

A. A Proposed Framework for Future Evaluation Efforts

It is recommended that AID/W abandon the Generic Scope of Work as the main approach to evaluating P.L. 480 Title II. Likewise, it is suggested that ad hoc evaluation efforts initiated by the various Bureaus and Offices be consolidated into an agency-wide comprehensive framework designed to meet their multiple information needs.

Specifically, the evaluation of feeding programs requires a multi-faceted approach. This may be conceptualized as follows.

1. Basic Research

First, there is a need to perform the basic research which provides the knowledge that serves as the foundation for the design of interventions. Fundamental questions such as the dietary requirements of malnourished infants; the relationship between maternal and child malnutrition, growth failure, and mental development; and the physiological response in terms of catch-up growth potential of a malnourished child, fall in the domain of basic research. This may be performed in a research facility or a field laboratory. Regardless, the concern is to expand the boundaries of knowledge which has generalizable application.

There are at present a number of basic tenets of feeding programs which require further research. Questions persist in the scientific community which make the design of programs a hazardous and precarious task. Likewise, other unknowns result in its being extremely difficult to predict program outcomes, or delineate expected achievements (e.g., the effects of incremental caloric intake in

maternal diets on birth weights of their children, and the constellation of factors which affect such outcomes). The lack of reasonable expectations makes it tenuous to design programs and thereafter measure impact in terms of poorly understood phenomena.

While pursuit of fundamental truths and basic knowledge is not the appropriate role for Food-for-Peace/W, it is recommended that they work to identify untested hypotheses and persistent questions which may be addressed by research institutions. This will provide long-term dividends to the Agency. A dialogue between researchers and practitioners will also provide useful guidance to academics and scientists in their endeavors, and help assure the applicability of the labors. Such an effort is best pursued in conjunction with technical offices, such as the Office of Nutrition and the Office of Agriculture. However, great selectivity should be practiced in delving into the research arena. In addition, efforts should initially focus on availing the Bureau and implementing agencies of what is already known, and on the careful design of programs that correspond to the state-of-knowledge.

2. Evaluative Research

Evaluative research is concerned with the application of existing knowledge to field settings in the form of interventions. As described by this author previously (Sahn and Pestronk, 1981):

Emphasis is placed on not only determining the benefit accrued to an individual from the manipulation of an independent variable in the environment, but also on assessing the improvement in the welfare of the community in general. As such, the feasibility of intervention in a community context is examined.

The purpose of an evaluative research activity, like all research, is rarely defined in terms of improving the welfare of the study community. Instead, the findings

are mainly utilized in the process of decision making and policy formation, since the focus of interest is on the viability of a given intervention strategy, and its replicability in other settings. Therefore, such evaluative research efforts often encompass such activities termed "demonstration," "pilot," "model" or "prototype" projects, which are being analyzed for their potential application elsewhere.

A major thrust of AID/W evaluation efforts should fall into this category. The insights gained from this type of evaluative research form the basis for determining program design and implementation strategy. These efforts must be characterized by their comprehensiveness and impeccable research design. Otherwise, findings will be ambiguous, and their utility compromised.

The intent here is simple: to make a determination of what outcomes can be expected in large-scale programs. If the management of inputs result in planned outputs, there remain questions concerning the linkages between outputs and purposes, and purposes and goals. So we look at operational field experiences, and extricate generalized principles to be used in the interactive functions of program planning, implementation, and monitoring and evaluation.

Throughout this report, areas for further study which fall in this domain are identified. For example, what are the expected long-term effects of various types of food-for-work projects on employment, income distribution, and economic growth? The answer, in conjunction with clearly stated objectives, will determine the nature of the FFW project undertaken. Another question is, whether the marginal propensity to consume food out of a transfer in the form of food is greater than in the form of cash. That is, do tied subsidies in the form of food result in greater food consumption, and if so, is it attributable to program characteristics such as nutrition education? Answers to these questions will, among other things:

dictate the size of the ration needed to show measurable improvements in nutritional status; indicate whether education is an indispensable component; show whether an income transfer in the form of food aid has greater nutritional impacts than a cash subsidy; and whether efforts to target programs should be based on income or nutritional criteria.

It is recommended that a systematic effort be made to answer these types of questions. In the past, they have been dealt with on an ad hoc basis. This is unsatisfactory. The answers provided are usually suspect, and many of the most difficult areas of inquiry are neglected. Representatives of the voluntary agencies, AID/W, and other experts should jointly guide such efforts. A beginning towards this end was made, for example, with the CARE Phase II and Phase III projects. Despite serious problems which are extant in research design, their intent is sound and these studies have and will continue to make a substantive contribution.

Concurrent with doing evaluative research to gather information about principles of programming, this type of rigorous experimental design can be employed by the Agency to determine impact of specific programs. This information may be required to guide the allocation of resources, or meet other purposes such as fulfilling the Congressional mandate. But like all field experimentation, careful planning is needed to provide clear results. And, of course, evidence that the major questions concerning process variables (as described in Chapter II) have been addressed, is prerequisite. In those few instances where field data exist and lend themselves to analysis, a retrospective design is feasible. But these instances are rare. Prospective field studies are preferable in almost all instances.

3. Monitoring/Evaluation Systems

The development of monitoring/evaluation systems in all Title II programs should be a goal of highest priority for the Agency. These systems are easily overburdened; therefore, the minimum amount of data to answer specific questions about program operations and impact should be collected. The major function of these systems is to assure services are being delivered, and that anticipated impacts are forthcoming. If not, further attention is required, in the form of scrutinizing program design and implementation.

It is emphasized that by the time a program reaches the stage of large-scale implementation, its design should be tried and tested. The fundamental hypotheses and assumptions on which the program is based should have been already examined in an evaluative research context. Thus, a monitoring/evaluation system identifies breakdowns in the implementation process, or unanticipated obstacles to reaching objectives.

Two other specific recommendations are germane to this discussion:

- Introduction of monitoring/evaluation systems should be carried out in a gradual and sequential manner. Initially, a few sites should be chosen and thereafter added to in accordance with the availability of trained personnel and the capacity of the new data collection system.
- Feeding centers which register an unusual level of program success, or of failure, should be given special attention and in-depth analysis. Much can be learned by comparing and contrasting extremes as to the ingredients which facilitate or impede the achievement of program purposes.

4. Process Evaluations

The examination of program processes and intermediate indicators covers a vast range of activities. The three subsets of "evaluation" described above are all concerned with program processes, although to different extents. But besides the concerns of research, evaluative research and monitoring/evaluation systems with program processes, it is legitimate to perform evaluations which concentrate on examining the diagnosis/assessment of the problem, the definition of objectives, the planning process, the administrating and implementing organizational structures, as well as the delivery of services. All of these steps fall in the domain of process evaluation.

When information about these areas is needed by AID/W, or when AID/W, USAIDs, or voluntary agencies suspect that there is a constraint to program success which results from factors such as confused objectives, sporadic service delivery, logistical hold-ups, and so on, it is legitimate and wise to dispatch evaluation teams to address such concerns. Much of the Title II evaluation effort to date has done just this, largely to fill an information shortage in Washington about field level programming. Two qualifications to continuing such assessments are: that it be clear as to what type of information is needed for decision making in Washington; but, more importantly, that the findings, especially those which identify impediments to effective programming, be part of a learning process for implementing agencies. Mechanisms to act rapidly in terms of improving design and implementation are essential.

B. Other General Recommendations

1. The above comments on process evaluation are a reminder that besides facilitating decision making, evaluation should be a formative exercise. A preliminary analysis of the findings of

evaluations already performed, coupled with the world-wide literature, is suggestive of the need to improve on planning and implementation as a concomitant to evaluation. Therefore, it is imperative that we use the vast knowledge from previous studies, and conceptualize evaluation as a way to improve programs, rather than resting on our laurels of self-criticism and accountability. Evaluation is easy and painless compared with making subsequent changes which stem from findings.

2. Evaluations, especially those examining process, should not only make recommendations about program improvements, but are advised to address explicitly how they are to be integrated into present operations. Information on monetary costs, management, personnel needs and so on should be included.
3. Before undertaking any type of evaluation activity, the purposes for doing so should be clearly defined. Thereafter, the appropriate "type" of evaluation should be selected to meet this stated function. Similarly, the specific information needs to be met by the evaluation should be made explicit.
4. Impact evaluations should only be performed or considered for projects where there is reasonable substantiation that the three themes of process evaluation, project context, program objectives and plans, and service delivery, have been addressed. That is, information that problems have been appropriately diagnosed, objectives clarified, plans and program design reflect objectives, and implementation takes place as intended, is a prelude to or concomitant with impact evaluation.
5. Measurement of impact must be done in terms of objectives, as delineated at the purpose or goal level. It is therefore important for evaluations to have clearly defined the program

objectives that they are attempting to measure achievement of, and that such objectives are consistent with those which the program was designed to achieve.

6. Evaluation teams should explicitly delineate the logical framework for projects being examined. This will serve to allow comparative analyses of the programs, make clear the expected impacts, and provide a means for reconciling different perceptions of program purposes among different actors and hierarchical levels of management.
7. Impact evaluation design should be developed to minimize competing explanations for findings and threats to evaluation validity. Given the impossibility of eliminating all such perils of evaluation, the more significant problems and biases should be explicitly addressed.
8. There is a need for interfacing with project functionaries and communities in order to explain findings which are often ambiguous. This suggests the adoption of iterative experimentation in the field.

C. Nutritional Impact

The evaluations reviewed, as with the world-wide literature, were generally not successful in quantifying nutritional impact. This is attributable to two factors: the inherent difficulty in measuring nutritional changes and attributing them to programs; and ill-conceived evaluation design. Despite the fact that most teams made qualified statements about impact, and presented data in support of their assertions, assessments appeared to be founded on process measures (intermediate indicators) and subjective opinions and observation.

Therefore, the following are recommended:

1. Problems inherent in evaluating nutrition impact do not require abandoning such efforts, but admonish against hastily developed, retrospective evaluation designs. They represent a squandering of resources, and do not provide sufficiently reliable findings from which to make programming decisions.
2. Given the perils and pitfalls of measuring nutritional status in the field, special care should be given to the selection of measures, the collection of data, and thereafter the techniques used in the analysis of data. This requires highly trained individuals in research design and nutrition program evaluation.
3. It is imperative to distinguish between changes in nutritional status which are attributable to a program, and those which occur due to natural phenomena observed in communities with endemic malnutrition. Specifically, two prominent and predictable factors deserve exceptional attention. The first is the phenomenon of how, in accordance with changes in the age distribution of a population, there are anticipated changes in the level of the malnourishment rate. The second is that due to statistical properties combined with the characteristics of nutritional status measures, the most seriously malnourished children will show the greatest improvement due to an intervention. And in the absence of an intervention, one will likely witness a spontaneous and expected improvement in those who are most seriously afflicted.
4. Two major approaches to measuring nutritional impact should be considered. The first is a series of carefully planned impact studies designed to test pre-determined hypotheses, using

pre-determined analytic techniques. These fall in the domain of evaluative research and should be performed on a select basis and commend themselves to the use of a prospective study design. Explicit understanding as towards what ends the finding will be used, and the types of decision alternatives which follow, is an essential component.

5. The second type of impact evaluation to be performed involves the emplacement of monitoring/evaluation systems to provide continuous and verifiable data on impact. Numerous advantages to this type of impact evaluation motivate its adoption. Most important is its propensity for making evaluation a formative process, in which improved program planning and implementation occur via established feedback mechanisms.
6. Nutritional impact should be measured using like criteria. This suggests the use of identical standards (i.e., WHO or NCHS), cut-off points, and indicators (e.g., weight-for-age). This will enable comparative analysis of future findings and make data less subject to the vagaries of the analyst.

D. Food-for-Work

The dominant problem in evaluating Food-for-Work (FFW) projects is the identification of the major objectives. Basically, FFW can be conceptualized as meeting two purposes: income subsidies via the creation of short-term employment; economic development via the creation of assets, infrastructure, and skills. While not antithetical, these two purposes usually conflict. Present FFW projects have generally tried to serve both the welfare and economic objective function, which has resulted in less than optimal use of resources. The evaluations have likewise viewed projects in terms of both objectives, providing an unfair and incorrect assessment of efforts.

Therefore, the following approach to evaluating FFW is recommended.

1. Programs emphasizing short-term employment and consumption goals (Type A) should be distinguished from those with long-term economic development goals (Type B). This may be difficult, as much confusion apparently exists in the formulation of objectives, and the design of projects which are responsive to those ends. Therefore, objective clarification and a reorienting of projects is prerequisite to evaluation.
2. The impact of Type A programs, which are primarily concerned with relief operations and meeting shortfalls in food availability, is measured in terms of whether rations are provided to those in greatest need. The concern is whether the income effects of the project are sufficient to forestall the inadequacies in consumption of food and other commodities which would otherwise jeopardize an individual's or community's welfare. Therefore, impact evaluation takes place during the "construction phase" of projects, when food is being distributed. This can be done through an on-going monitoring/evaluation system at all project sites.
3. Type B programs are primarily to be evaluated in terms of their long-term stream of economic returns which accrue to the asset or infrastructure created. These returns are measured during the "operational phase" of FFW, when the food is no longer being distributed. A number of parameters exist for measuring benefits. They include: effects on agricultural production, structural employment, and income distribution. While these impacts may prove elusive to measure, proxies such as land value, flood protection, access to markets, the degree of migration, and so forth, can be substituted. Measuring impacts of economic development projects must be done in a limited

number of project sites. This type of "evaluative research" requires sophisticated experimental design. The thrust should be to uncover generalized principles characteristic of project "types."

4. Although impact of Type B programs is limited to select projects, process measurement of all projects is paramount. The collection of intermediate indicators such as whether inputs were provided (e.g., food, tools, materials), an assessment of labor productivity, and verification that the asset of infrastructure was completed, is mandatory. These data should be collected as part of an on-going monitoring/evaluation system.
5. Improved nutrition should not be an impact measure for FFW.
6. Evaluations focused on process should give increased attention to how well FFW projects are integrated in Mission CDSSs and a country's overall development priorities and strategies. They should assess to what degree food aid in general and FFW projects in specific are complementing other USAID activities in areas such as agriculture (e.g., irrigation, flood protection), rural development (e.g., flood control, reforestation), and women in development (e.g., training).

E. Educational Objectives

The evaluations reviewed described two groups of objectives for school feeding: educational and nutritional. Some programs stated one or the other as their purpose, while others designated both. It is felt that program designs should differ according to objectives. Likewise, impact evaluation should only be performed in reference to achievement of program purposes. Therefore, a clear statement

of desired outcomes is prerequisite to further study. The lack thereof was an impediment to the evaluations reviewed. Program design and implementation also suffer.

The following recommendations are presented not only to improve the evaluation process, but also its complements, the planning and implementation procedures.

1. Records on enrollment, dropouts, and absenteeism should be regarded as satisfactory proxies for measuring impact. In order to determine whether food acts as an incentive, a sample of feeding and non-feeding schools should be selected, representative of different environments and operational contexts. Thereafter, an ongoing monitoring/evaluation system should be initiated to track the effects of the program on the proxy measures. This procedure is reasonably facile, since most schools presently maintain records. Therefore, the major effort would involve setting up a management information system.
2. Consultation with experts is suggested to design the most feasible ways to measure more directly program impacts using parameters such as school performance and academic achievement. These studies should be considered on a limited evaluative research basis. They will provide further substantiation for the hypotheses which link outputs (school feeding) to purposes (improved education).
3. It is not recommended that AID sponsor any further research concerning the linkage between improved education and the goal of economic development. This is already well documented in the literature.
4. It will be difficult to measure nutritional impact of school

feeding programs. It will be even more unlikely that impacts will be as encouraging as MCH nutrition projects. Justifying programs on the basis of nutritional impact is precarious and will probably result in the elimination of the program. Therefore, if social and economic development are priorities or higher program goals, nutrition objectives should be dropped and education objectives embraced.

F. Economic Impacts

1. Attempts should not be made to measure the impact of Title II programs on national macroeconomic aggregates.
2. Measurement of Title II effects on regional economic aggregates (e.g., inflation, agricultural production) and the microeconomic production decisions of the household should be confined to FFW projects. This should be done on a limited evaluative research basis.
3. The types of questions to be answered by microeconomic analysis of MCH and SF programs are the determination of: (1) how nutritional benefits accrue to the family via income effects; (2) to what extent a tied subsidy in the form of food is "more beneficial" than a cash transfer, as measured by its effect on individual and/or household food consumption; and (3) what consumption changes other than increases in the food budget are manifest by the family, and to what extent they are considered beneficial impacts. Answers to all these questions must be based on household income and expenditure surveys. Thereafter, they will allow for the empirical determination of the marginal propensity to consume out of income versus in-kind payment, the elasticity of food demand versus caloric demand with respect to income, the valuation of the food ration, and so on. It must

be recognized that estimating these parameters requires sophisticated and expensive field research. It follows that microeconomic analysis should not be undertaken on a widespread scale. Specific research projects may be appropriate to substantiate the conceptual basis for programming. But, in general, program impact should not be measured in terms of microeconomic parameters.

4. Most of the evaluations reviewed discuss income effects. There is a need to develop a standardized methodology to value the ration. The approaches used to date have been inconsistent and, in many cases, flawed.
5. Program objectives can be stated in terms of (a) nutritional impacts or (b) income effects. In the former case, impact is measured by improved nutritional status, and achievement is via the ability to target the ration as well as the general in-kind income subsidy. In the latter, it is sufficient to document that the income transfer took place, and poverty lessened. However, it is strongly advised that programs are not justified on the basis of income effects, since food aid is not a cost-effective way of transferring resources. Therefore, it follows that it is inappropriate to measure income effects as an indicator of program impacts.
6. Evaluators should distinguish between programs that have "economic effects" (e.g., resource transfer) and programs which promote "economic development." The latter is measured by self-sustaining growth of income, economic product, and/or more equitable income distribution, which is likewise not a transient (static) phenomenon.

G. Integrated Programs and Related Impacts

1. Nutrition education, sanitation, health care, and related complementary components of feeding programs are intermediate indicators. Their measurement is not to be mistaken for impact evaluation, which involves the measurement of nutritional status.
2. Experimentation which determines to what extent integrated outputs (e.g., feeding and education) act synergistically or additively to improve nutritional impact is an evaluative research question demanding further field study.
3. It would be overly cumbersome to distinguish the impacts of different components in large-scale interventions. Therefore, a composite indicator, such as nutritional status which is a function of a number of different services, should be used in on-going monitoring/evaluation systems. However, the monitoring aspect should be focused on the assurance that the distinct services are being provided as planned.
4. A series of tangential effects or objectives of Title II programs are often highlighted by evaluators as criteria for measuring success. These include assisting people in learning to help themselves, developing a sense of community self-reliance, providing a mechanism for socialization of children and mothers, and so forth. It does not seem to be a propitious use of resources to evaluate these fundamentals. Rather, assurance should be provided that they are built into the project and are central philosophical tenets.

20

III. PROCESS EVALUATION

The emphasis of this paper is on the determination of the impact of P.L. 480 Title II programs. Nevertheless, I feel it imperative to heed the axiom of "process before impact," and include the following brief discussion on the role of process evaluation. It is emphasized that until the steps which fall in the domain of process evaluation have been explicitly or implicitly addressed, it is premature to consider impact evaluation.

Another reason for including a chapter on process evaluation is that in reviewing evaluations of P.L. 480 Title II programs, especially those which have been performed under the Generic Scope of Work, one cannot help but be impressed with the amount of interesting and informative documentation amassed by teams during a few weeks in the field. However, the vast majority of their efforts fall within the realm of process evaluation. The range of topics covered includes such diverse areas as the economic setting; the nutritional setting; logistical issues of moving and accounting for the commodities; detailed descriptions of operational procedure; illustrations of feeding site activities; determination of objectives; delineation of program costs; examination of administrative structures; and so on.

After reading the evaluations from Ghana, Kenya, and Upper Volta that followed the Generic Scope, and others such as those from India and Egypt, one is impressed that they effectively focused on many constraints to program impacts. Such an exercise is valuable in identifying points where breakdowns occur and program effectiveness is hindered. When such findings are understood and acknowledged by program functionaries, the evaluation serves as a learning process and will enable implementors to concentrate on improving operations. It is questionable, however, whether this last purpose was met, given the

short time frame of the efforts and the problems inherent in an outside evaluation.

In a similar vein, an examination of program processes, plans, and country-specific problems and issues proves a valuable service for AID/W. It provides them with an opportunity to extend their ears and eyes into the heart of program operations. As such, it serves as an information link between in-country programs and AID/W, and makes the center of decision-making aware of the realities and hardships of those out in the field.

Given the diversity of purposes for and aspects of process evaluation, it is advised that the task be conceptualized in three sequential steps. These are the subject of the discussion that follows. (See Sahn and Pestronk, 1981, for a more complete discussion of the theoretical basis of the following discussion.)

A. Project Context

The component of process evaluation is concerned with:

- pre-existing indicators which describe the environment in which the program is set, focusing on the social and economic context; and
- organizational and administrative structures of implementing agencies, looking at funding procedures, decision-making roles, and so on.

The evaluations reviewed gave adequate attention to examining the project context. All provided information on the nutrition setting. The organizational structure of the implementing agencies, relationships of programs to governmental efforts, and the economic climate in the countries examined were also broached. Some evaluations, however,

spent an inordinate amount of energy providing what is ostensibly background information, which is to be used to frame the analysis of the program.

In general, the provision of contextual information to AID/W does not necessitate dispatching teams to another country to gather information on the nature and extent of nutritional and related problems in a country. A great many sources of country-specific nutritional status data are available; likewise, the socio-demographic and economic information which makes up a large portion of a number of the Title II evaluation reports is readily available from a variety of other AID, World Bank, and other sources. Therefore, it is stressed that the provision of such contextual data is not evaluation, in and of itself. Its value is as an introduction to the reader. It provides benchmark data to diagnose problems and judge whether a programmatic response is warranted.

B. Program Objectives and Plans

The second major aspect of process evaluation combines the exploration of program objectives and the appropriateness of the response in terms of program plans. Questions are asked such as: Is nutritional improvement an appropriate objective for food aid programs? or Are there more worthwhile purposes which can be achieved with the resources? If there is a food-for-work program, one may ask if it is designed to address an underlying constraint to development. And, of course, it must be determined whether plans are realizable given administrative capacity and infrastructure limitations in a given country.

To illustrate this aspect of process evaluation with a concrete example, let us examine the question concerning the appropriate ration size for a program. There is a large variance in the nutrient value of rations in MCH and SF projects. No clear-cut criteria

exists to evaluate what the correct ration size should be. Rather, we must carefully consider the design parameters of a project, which include ration size, and how they relate to objectives.

Specifically, if a program is hypothesized and designed to improve nutritional and health status through the direct provision of a supplement, it will undoubtedly necessitate a high caloric ration. Reasons include that the malnourished children have a large caloric gap between their needs and intake, and that the ration must compensate for leakages in the form of substitution and sharing. On the other hand, if it is hypothesized that a program's nutritional impact is mainly through indirect effects, where the ration serves as incentive for participation in allied health and educational activities, the aim is to provide the smallest possible ration which will encourage regular attendance at the health center. This will allow for maximizing population coverage.

Similarly, a school feeding program with educational objectives should attempt to provide the smallest quantity ration in order to attract students to school. Conversely, if the primary objective is nutritional, a large caloric increment is required to compensate for leakages and the limited number of school days per year.

Essentially, the purpose of this aspect of evaluation is to determine whether programs are designed to return the most benefit in terms of meeting objectives which respond to real needs. Therefore, falling under this category of process evaluation is reviewing the appropriateness of program designs as presented in voluntary agency Program Plans, CDSSs, and other documentation. Ideally, if sufficient attention were given to formulating objectives, and program planning and design activities, this step in the constellation of evaluation activities would be perfunctory and brief. But given the reality that food-aid programs have historically been hastily conceived, without proper

Attention to program design and implementation, the focusing of evaluation activities in this realm is valuable.

In the future it might be more helpful to label such activities as reviews or assessment, as was done in the case of the Egypt report. But more importantly, the concept of an "outside evaluation" often provokes uneasiness or even hostility in those being scrutinized. Therefore, criticisms are often not understood or heeded. Self-defensiveness among field workers is characteristic. Thus constructive recommendations may be overlooked or dismissed.

Therefore, when teams are sent out by AID/W to fill an information shortage, or scrutinize program objectives and analyze their operations, emphasis should be placed on assisting the voluntary agencies and their counterparts to improve the program. This will necessitate a different type of technical assistance than used in impact evaluation. And to expect a team to do both within a few short weeks in the field, as had been the case to date, is unreasonable.

C. Service Delivery

Substantiating that services were provided as planned is the thrust of this component of process evaluation. Documentation that a program reached its target group (or that inputs were transformed into outputs) and services delivered are intermediate indicators of impact. These intermediate indicators are more easily measured than impact measures. If process evaluation indicates that services are not provided, e.g., that food and other inputs are not available and implementation does not take place as planned, there is no reason to search any further for impact. Instead, the task is to figure out why and make the appropriate adjustments. If activities appear to be going as planned, evaluation of impact is appropriate.

The above comments would seem self-evident. But, in practice, they are often overlooked. Witness, for example, the Ghana evaluation. A good portion of this document is devoted to recounting the serious obstacles to logistics and program operations. Hopefully, AID/W was previously aware of most of these difficulties and the team was simply verifying what was known, providing an update on the situation. Therefore, one must initially ask whether the team performed a vital function in that regard. But more significantly, it is a curiosity that concurrent with a disparaging assessment of the program operations, the evaluation attempted to collect and analyze data on program impacts.

To amplify, consider a variety of facts presented by the team:

- (1) ". . . there is no screening process for children in rural areas."
- (2) ". . . there was little consistency in the amount or types of commodities provided to participants . . ."
- (3) "there was a lack of consistency in receiving the program allocations of food."
- (4) ". . . the variance in attendance for individual participants is extreme."
- (5) "Beyond the lack of consistency in food distribution, the major difficulty in determining program impact related to commodity consumption, is the impossibility of determining the actual consumption by recipients."

In total, such observations preclude performing impact evaluation. They admonish implementing agencies and the Mission that programs are in disarray. I presume that the teams' insight was not startling to anyone, especially in Ghana. But regardless, the message is clear: to upgrade the feeding programs, either by altering design, or by concentrating on new techniques of service delivery. That should be the charge of any teams sent out to Ghana in the foreseeable future.

While Ghana is an overt illustration of the need to assess process

before impact, similar scenarios could be outlined in regards to other evaluations. There are serious planning and operational problems with many programs. This demands that attention be paid to those issues, and that AID/W make the appropriate investments to rectify the situation. It is easy to identify program shortcomings and failures. The creative and demanding task is to bring about the changes which assure or at least enhance the possibility for success. Such efforts should be a necessary complement of any evaluation.

27

IV. NUTRITIONAL IMPACT

The use of food aid as a resource has been inextricably linked to nutrition. The most obvious result of this equation is that most Title II programs state nutritional improvement as the, or one of the, major objectives. It is therefore incumbent upon the evaluator to measure program success in terms of impact on nutritional status of its beneficiaries. A critical examination of how evaluation teams have done just that, and the implications for future evaluations are the focus of this discussion.

A review of the attempts to determine nutritional impact of Title II maternal and child health and school feeding programs suggests that they made heroic efforts to provide quantifiable findings. Unfortunately, the difficulties inherent in evaluating nutritional impact have all but precluded placing credence in the quantification which has been attempted. This assertion suggests the need for carefully considering the difficulties encountered while attempting to quantify nutritional impact. Furthermore, if the mandate is to quantify program effects, alternative approaches must be considered, which will also be the subject of this discussion.

A. Previous Evaluative Efforts

The evaluations reviewed determined nutritional status by developing procedures to analyze data that were readily available in the field, or sufficiently simple to generate within severe time and resource constraints. More often than not, the available data did not lend themselves to being retrofitted into a hastily conceived and much-compromised evaluative design. That is, when there were data available in the field, or if the team was able to collect limited data during their short field exercises, they were rarely sufficient

or appropriate to become a part of an evaluative framework which would lend believability to quantitative findings.

The result was that ad hoc observation, subjective judgments, preconceived notions, and expert opinions were more often relied upon than not in the quest for proving impact. This is not to demean or invalidate such an approach. In fact, expert judgments and situational assessment make up the vast majority of food and nutrition evaluations. (For example, World Food Programme Interim Evaluations, which involve dispatching a team of experts into the field, fit this mold.) And even when quantitative data are available, their interpretation is subject to the perceptions and feelings of the analyst. However, reliable data generation systems can go a long way to substantiate or verify what would otherwise be subjective judgments about nutritional impact.

This problem of futile attempts to quantify impact has plagued the evaluations done under the Generic Scope of Work, as well as most of the Philippines impact evaluation. Similar difficulties, although to a much lesser extent, are manifest in the Morocco evaluation, whose expressed purpose was to examine impact. By concentrating considerable energies just on the question of impact, these two evaluations were able to muster resources to present convincing findings. In contrast, the limited time frame of evaluations done in Upper Volta, Kenya, and Ghana, and the numerous other demands on the evaluation teams in terms of examining program processes, coupled with the unavailability of existing data, make it unreasonable to expect the teams to have substantiated impact.

While I could critique individual evaluations of Title II programs, it will be more constructive to recount briefly consistent methodological problems. Occasionally, these will be illustrated with country examples.

1. Measurement Problems

A variety of indicators has been employed for measuring nutritional status. Anthropometry is usually relied upon in field settings. Nevertheless, a myriad of problems exists, which makes findings of evaluations based on such indicators suspect. Before designing any further evaluations, the reader is at a minimum urged to refer to Habicht (1980), Martorell, et al. (1980), Miller et al. (1977), and Sahn and Pestrunk (1980) for discussions of selecting indicators. Although this is not the forum for an extensive discussion of such problems, a few salient points deserve special mention. The intent is to emphasize that measurement of nutritional status is far from a science.

a. Collecting Accurate Data

Initially, there are problems of accurately collecting data in the field and thereafter transcribing data into written form. All of us who have attempted to gather anthropometric data in the field know what a perilous task it is; and even the astute observer of village-level workers measuring and recording weights, overburdened and often not convinced of the utility of such a cumbersome task, will quickly realize the propensity for inaccuracies. The evaluations reviewed generally recognized this problem. Furthermore, they noted time and time again that data were often missing, unavailable at certain centers and potentially inaccurate for reasons such as inadequate measuring devices and techniques. More often than not, inquiry into quantifiable nutritional impact should stop at such a point.

While the difficulties of collecting field measurements are not

a new theme, it is often acknowledged and brushed aside as a recurrent problem which due to its pervasiveness can be ignored. This is felt to be the wrong approach, as measurement errors do not cancel one another out and can have serious effects on program findings.

To illustrate this point, and inspired by the findings of the Upper Volta report, we ran an experiment to make more tangible the sensitivity of results to the smallest biases introduced by a scale not zeroed, or a health worker rounding results up or down to the nearest half or quarter kilogram.

Specifically, the draft report on an evaluation of PL 480 Title II programs in Upper Volta claims that measurements taken in successive Januarys in 13 schools suggest a positive effect of school feeding programs. Heights and weights were taken for the school children in January 1980 by school officials and in January 1981 by the evaluation team. Due to a delayed opening of the schools in 1981, children had not been receiving food supplements prior to being measured. In contrast, in 1980 the children had been receiving school lunch for three months. Overall, the percent of the children deficient in weight for height (defined as less than 90 percent of an unspecified standard) in 1980 was 31.4, while in 1981 the percent increased to 43.6.

As an experiment, we considered the weights and heights of all children between the ages of 73 and 96 months of age as recorded during the Rice Fortification experiment in Thailand and archived at CSF/Ann Arbor. The data were collected in January 1973. Using the NCHS-CDC standard (also called the WHO standard), the percent deficient (also defined as less than 90 percent) was computed on the Thai data set. We then

added 0.3 kilograms to each weight and recomputed the percent of normal and deficient. The addition of 0.3 kilograms to each weight resulted in a decrease of the percent deficient from 36.9 to 25.2--a difference of almost the exact magnitude as the difference observed in Upper Volta.

What does this mean? It does not suggest that the findings of the Upper Volta study were invalidated due to poor measuring techniques. It does illustrate that when using anthropometric measurements, especially those taken under difficult field conditions, very small idiosyncracies in measurement devices can have far-reaching consequences on analytic results. Anyone who has tried to maintain a scale used repeatedly in the field to a true zero knows how easily a consistent 0.3 kilogram error can be made on a given day (or in a given month). This is all to say nothing of the fact that the weight of an individual may vary by a kilogram depending on the time of day the weight is taken, or whether shoes are worn or not! As such, consistent measuring technique is fundamental.

b. Inconsistency of Measures

A related problem is the propensity for misclassifying children's nutritional status when anthropometric measures are used. When weight data are used, it must be recalled that there is great variability in the genetic growth potential of individuals, as a broad spectrum of weight is normally distributed among the population. Therefore, a child's position on the weight chart is much more a function of genetic potential than nutritional status. In addition, weight for age, which is a composite of stunting and wasting, may be low due to deficits incurred years previously and not to present status. These children would be misclassified as malnourished. Conversely,

an edematous child may be classified as normally nourished, when in fact he has kwashiorkor.

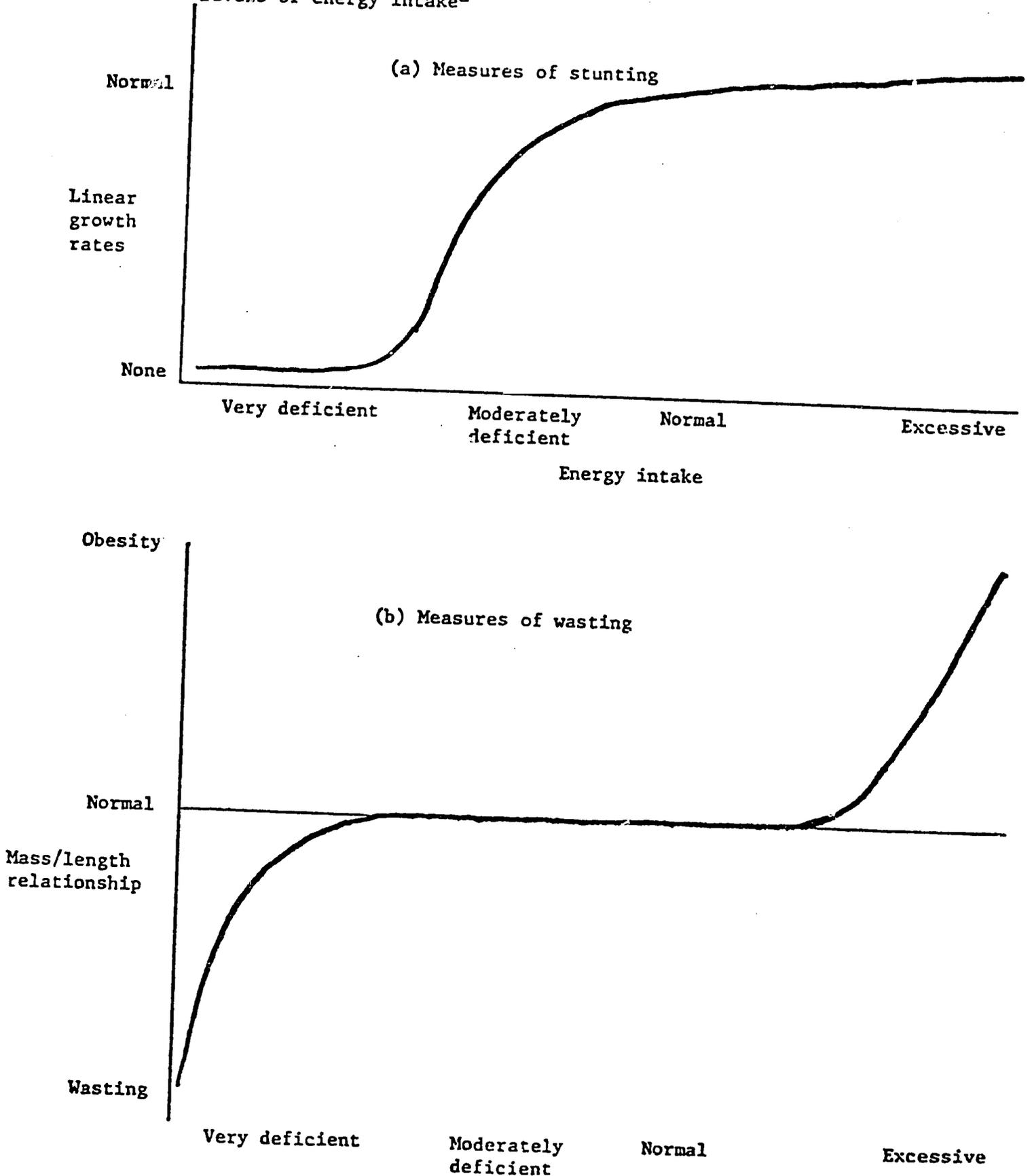
Similarly, although height-related measures are less prone to misclassifying malnourished children due to genetic variability or edema, height is insensitive to a child's present nutritional status. Height deficits take a long time to develop. In addition, the distribution of heights in a population is narrow; effects of small measurement errors are magnified. Heights are also more difficult to measure in the field and are therefore less precise.

In addition, it must be realized that height and weight indicators are not equally sensitive throughout the range of nutritional status. This is illustrated by Figure 1, which shows that, for example, when children have very deficient intake, linear growth rates such as height-for-age are of little use, while other measures such as weight-for-length are more appropriate.

Once again, to illustrate the vagaries of measures in correctly classifying malnourished children, consider the following scenarios based on our experience at CSF doing empirical analysis. A child is malnourished according to a weight-for-age classification, due to a previous episode of malnutrition. But the same child may well be normally nourished according to weight-for-height indicators. Thereafter, if the child goes through a rapid growth spurt in length where weight is gained and the child "catches up" some of his height deficit, you may find a flip-flop in indicators of malnourishment. Because the child stretched out and gained length so rapidly, his body proportions may change markedly. He may now be malnourished according to a weight-for-height standard, despite weight gain being

Figure 1

Response of anthropometric indicators to varying levels of energy intake^a



Source: Martorell, 1981.

sufficient to show improvement in nutritional status as a function of weight-for-age. Another example of the same type of phenomena is that you will almost always expect weight-for-height and height-for-age data to flip-flop over time, with one indicator showing improvement and the other deterioration. This is simply a result of the mathematical property that the height is the numerator in one measure and the denominator in the other.

The choice of measures clearly determines findings. Indeed, a composite index using weight and length measurements is preferred (see, for example, Waterlow, 1972), but may not be viable in the field. It is therefore important to compare data which used the same measures, and not jump to any conclusions on the basis of data collected over short intervals of time, especially if randomization is not used in selecting treatment and control groups.

A final point to be made is that although anthropometrics are universally accepted as the tested technique for measuring nutritional status, a child's response to feeding may not be manifest in height/weight measures. Explanations for this are that increased intake may result in a restoration of normal metabolic activity; there may be an increase in energy expenditure through greater physical activity; and that stunting at an early age may be irreversible, which would imply that it is impossible to regain height to the degree as to allow weight gain to be sufficient to show a change of status according to the anthropometric measures. A further problem with measuring impact relying upon anthropometric measures is the differential rates of maturation. This is especially acute for school feeding progress where children 10 years of age and older approach puberty and make measurement of nutritional

status extremely difficult.

2. Analytic Techniques

a. Standards and Cutoff Points

In addition to the problem inherent in measuring children, there is a great deal of leeway in selecting the analytic techniques with which to examine data once collected. Specifically, there is an array of growth standards and cutoff points for classifying children according to different degrees of malnourishment. The choice of which analytic tools to use is dependent upon a variety of factors, ranging from the country in or organization for which one is working, and personal preferences based on familiarity.

For example, in Ghana CRS employed a Growth Surveillance System in which the Harvard System was used as a base, and thereafter "substantial modifications were indeed made." In turn, the Ministry of Health's road to health chart was found to show a "considerable difference, sometimes about a kilogram," from the CRS charts. Thus, two very different standards were being used in the same country, which in turn were modifications of an internationally accepted standard.

Similarly, when this author reviewed evaluations of CARE, CRS, and other MCH feeding programs in India (Sahn, 1980), numerous classifications of Grade I, II, and III malnourished children were employed (e.g., in one evaluation, 80-89 percent of the mean was Grade I, and another instance, 76-85 percent of the mean received the same designation). Once again, the choice of cutoff points as with the choice of standards, often an arbitrary decision of the analyst, will result in different

findings of evaluative studies (see Table 1 as an illustration).

It is therefore recommended that AID/W encourage the adoption of a single standard (preferably those prepared by NCHS, also referred to as the WHO standard). This will result in more objective evaluations of country programs. But of greater importance is that if different standards and cutoff points are used in evaluating programs, there is no basis for comparing results, unless all raw data are subsequently re-analyzed after conversion to the same standards. This is a tedious task. So if AID/W has an interest in doing a comparative analysis of program findings, it is mandatory that the same indicators be used.

b. Statistical Significance

Concerning the issue of tests of significance and use of statistics in general, it is widely recognized that a great deal of license is left to the statistician and analyst in making a determination of the results. Once again, the reader is referred to a monograph prepared for AID, pp. 54-57 (Sahn and Pestronk, 1981) for an example from an MCH program which illustrates how such subjective judgements are in the eye of the beholder, rather than a function of the science of statistics.

Another example is noted in the draft of the Upper Volta school feeding report. In an attempt to determine nutritional impact of a school feeding program, the evaluation team was able to take advantage of the late opening of school in 1981, as described previously. The use of this comparison group design essentially matched the treatment group of 1980 school

Table 1

Nutritional Status of the Same Population, Using
Identical Weight-for-Age Data with Different
Standards and Cut-off Points

		Grade III	Grade II	Grade I	Normal	Total
GOMEZ	Classification (%)	30-60	61-75	76-85	86+	
	Number	6	85	210	526	827
	Percent of Total	.7	10.3	25.4	63.6	
HARVARD	Classification (%)	30-60	61-75	76-85	86+	
	Number	10	170	176	372	828
	Percent of Total	1.2	20.5	33.3	44.9	
GOMEZ	Classification (%)	30-69	70-79	80-89	90+	
	Number	34	134	223	426	827
	Percent of Total	4.1	16.2	28.2	51.5	
HARVARD	Classification (%)	30-69	70-79	80-89	90+	
	Number	72	224	257	275	828
	Percent of Total	8.7	27.1	31.0	33.2	

children with 1981 controls.

The results of the analysis of over 3000 children indicate that according to the aggregated weight-for-height data from the 13 schools, the school feeding program has a nutritional impact. The difference between the proportion malnourished in 1980 and 1981 was reported at $p < 0.001$ (see Table 2). Rather impressive--but wait. Upon disaggregating those data by schools, it is noted that in only ^{nine} ~~five~~ of the 13 schools, the percentage of children whose weight for height was less than 90 percent had deteriorated in 1981, when there was not a feeding program. Over one-third of the schools did not display any impact from the program. The disaggregated data therefore do not corroborate the "highly statistically significant" findings presented by the team.

The more interesting question is why in the Doure school did 11 percent of the children have weight-for-age less than 80 percent of the mean in 1980 and zero percent in 1981, while in Yako, one percent was less than 80 percent in 1980 and 10 percent in 1981. My contention is simply that if the evaluator cannot provide insight into such questions, the statistical results may be aberrant. The lack of a hypothetical model precludes giving serious consideration to such methodologically suspect social research. But of greater importance, the learning value from the findings, even if accurate, is compromised. Instructive lessons and valuable understanding would be gained from an effort to explain the conflicting results in two villages. Such exploration would provide a piece of the puzzle as to what circumstances or environmental factors enhance or impede greater growth and development of school children.

TABLE 2

Nutritional status of participants in the school lunch program: a comparison between early 1980 (after receiving lunch for 3 months) and early 1981 (after no school lunch for 3 months).

Schools	Jan. - Feb. 1980			January 1981		
	< 90 % wt/ht	< 80 % wt/ht	N	< 90 % wt/ht	< 80 % wt/ht	N
Ramengo	27 %	2 %	181	43 %	4 %	181
Poa	43 %	4 %	183	68 %	11 %	151
Koudougou (sud)	23 %	1 %	407	39 %	10 %	415
Koudougou (est)	29 %	2 %	414	21 %	3 %	449
Issouka	18 %	1 %	177	42 %	5 %	153
Bourkina	26 %	1 %	245	44 %	7 %	286
Arbollé	22 %	1 %	235	23 %	1 %	298
Douré	65 %	11 %	158	17 %	0 %	77
Yako (mixte)	19 %	1 %	450	48 %	10 %	387
Yako (protestant)	35 %	0 %	164	33 %	2 %	138
Banankyo	36 %	2 %	166	56 %	9 %	218
Goundi	34 %	2 %	130	16 %	1 %	183
Reo (mixte)	16 %	3 %	238	45 %	4 %	329
Total	29 %^a	2.4 %^a	3426	38 %	5.6 %	3265

^a The difference between proportion malnourished in 1980 and in 1981 is significant at $p < 0.001$.

3. The Difficulty of Attribution

Even if it is possible to substantiate a change in nutritional status of the community, there are a number of factors which impede the attribution of improvements in nutritional status to program activities. Such confounding factors can be negative or positive, in the former case making a successful intervention seem ineffectual, and in the latter an unsuccessful intervention appears to have improved nutritional status. Prominent among these confounding variables are those natural changes and phenomena which one observes in most populations with endemic levels of malnutrition. That is, there are predictable changes in the nutritional status of the populations that are expected regardless of whether there is or is not an intervention. In addition, secular changes in the project environment and other phenomena related to the delivery of services further aggravate the difficulties of relating changes to program activities. These are largely discussed below.

The most predictable competitor with a program as an explanation for nutritional changes in the population, which was overlooked in every study except the one done in Morocco, is the phenomenon of population aging. In virtually every data set we, at CSF, have seen (we have 10, or more, such data sets in our archive at CSF), the rate of malnourishment in the preschool population varies with the age of the children. Newborns are near normal. At about eight or 12 months of age, a large number of children ceases to grow (or slows up dramatically). At somewhere between 18 and 26 months of age, many of the children (now malnourished) begin to grow again. Even in the absence of intervention, many of the children grow at a pace slightly faster than that prescribed by most growth standards; therefore, in the aggregate, the rate of malnutrition goes down as the average age in the

18-24
-11-1-1968

2-5
1/2/15

population increases beyond the critical of maximum determination.

A second although less important phenomenon observed in malnourished populations concerns the contention that those with the worst nutrition status benefit most from a program. This statement was universal in the evaluations reviewed. However, the conclusion that the intervention had differential effects, helping those most severely malnourished to the greatest extent is questionable. Once again, naturally occurring changes in populations can once again explain such findings. Specifically, these are twofold.

In populations, there is a spontaneous movement of individuals who fall at extremes of a measurement scale (i.e., Grade III malnourished) to regress toward the mean (i.e., improved nutritional status). This is logical. If you have Grade III malnourished children, either they stay the same or improve (or alternatively die, which in most cases means they are no longer part of the sample). On the other hand, you would expect a certain number of Grade I malnourished children to improve and the remainder to worsen, becoming Grade II.

Compounding this problem is the realization that indicators such as weight-for-age are not equally sensitive at different degrees of malnutrition. Habicht (1980) points out that weight is an excellent indicator for measuring severe malnutrition, and increments in weight are extremely sensitive to changes in status. This is not so for the mildly or moderately malnourished, where the increments in weight as a result of improving diet are much less sensitive. (Just the opposite is true for height measures.) Thus it is erroneous to consider weight-for-age an equally good indicator for a wide range of malnutrition. This

implies that it is inappropriate to compare the benefits of a program between severely and mildly malnourished children on the basis of their degree of weight gain. Unfortunately, this was done in most evaluations.

Thus, in any data set drawn from a community experiencing serious malnutrition, we would expect:

- a steadily rising rate of malnutrition with age among the youngest children; and
- a steadily declining rate/ ^{of malnutrition} among the older children; and
- the greatest improvement among those children who were worst off but able to survive.

I had no infants in the sample

2-5 year olds - no natural decline in malnutrition basically stay same

Now let us take a look at one of the more ambitious uses of sophisticated statistical techniques from among the P.L. 480 Title II evaluations done recently--the regression analysis of MCH programs in the Philippines. (See Appendix B.) Assuming the validity of the sample, it is learned from the evaluation that:

- (a) 24.4 percent of the sample of 238 were infants (between six and 11 months of age);
- (b) nutritional status declined most frequently in this age group--59 percent as compared to only 12 percent in the balance of the population;
- (c) the percent normal or near normal was highest in this age group, 59 percent, with the second highest percentage--21 percent in the one- and two-year-olds.

This is in accordance with the classic patterns described above of relatively normal infants going through a period of deterioration.

The evidence that the balance of the pattern holds is more

difficult to demonstrate from the numbers presented in the report. However, by noting that ages at time of entry must have included a sizable proportion of children of the age at or beyond the point of peak deterioration, we can deduce from the correlation analysis that the classical pattern holds. There must have been more older entrants than younger ones because only 24 percent of the sample were infants and almost 60 percent of the remaining children in the sample had participated for less than a year. The positive correlation between age of entry and improvement in nutritional status therefore implies that the children beyond the point of deterioration were, in fact, improving.

Finally, we cite the high correlation between nutritional status at time of entry and improvement as evidence that the worst children do, in fact, improve most. (Note, this is almost a tautology--after all, children of nutrition level 1, 2, or 3 cannot improve much, while children at levels 7, 8, or 9 have a much greater potential for gain.)

We ask, then, what does the regression analysis tell us?

The equation is given as

$$DNL = -3.317 + 0.483(ENTNL) + 0.309(ENTAGE) + 0.554(LOP);$$

where DNL = change in nutrition level (1 is best, 10 is worst),
 ENTNL = nutrition level at time of entry,
 ENTAGE = age at time of entry, and
 LOP = length of time in program.

It too confirms that the classical pattern exists. We would expect, even in the absence of intervention, that the change in nutrition level would be related to the chosen explanatory variables as in the equation. The worse the beginning nutritional

level, the greater the improvement; and the higher the age of entry (and most children enter beyond the age of peak deterioration), the less the improvement. (This last relationship demonstrates that the nearer to the point of maximum decline a child enters the program, the greater his improvement.)

These are expected results and not proof of program impact. The only variable in the equation which might shed some light on impact is length of time in the program. Although this variable is statistically significant, its contribution to explaining change in nutritional status is almost negligible. Does the program work?

The authors seem to think that because the sample drawn behaves precisely as one would predict--even in the absence of intervention--that the program is successful, even though the only program variable in the equation has hardly any effect. The talk of statistical significance masks the obvious--beyond a certain age children improve and, therefore, a cleverly selected sample is bound to show improvement.

We use the Philippines data to illustrate this point because it used the most sophisticated statistical analysis. A similar critical analysis was done for other evaluations as well.

Take, for another example, the Ghana evaluation. They conclude on the basis of Table 3 that the "data indicate that the worse the level of malnourishment the progressively larger mean weight gain were found." First, the mean weight gain varied by 0.15 kilograms for different levels of malnourishment. Given the small sample sizes and the lack of precision in collecting weight data, this finding must be viewed as insignificant. Nevertheless, even if an argument could be made that the data were absolutely

Table 3

Six Months' Weight Gains in Kilograms of
Sample Ghanain Children by Growth
Surveillance Percentile

Percentile/ Diagnosis	Number of Children	Total Gain	Mean Gain
80 - 100 +	132	209.6	1.58
60 - 79	94	154.3	1.64
59 and below	15	26.0	1.73
Kwashiorkor/ Marasmus	31	60.4	2.20

Source: Ghana Draft Report.

accurate, these findings would be expected given the tendency to regress toward the mean and the fact that for the most malnourished children, weight-for-age is a more sensitive measurement. It is therefore hard to argue that the most malnourished benefited most from the program. Furthermore, the team's suggestion that the "data showing progressively greater mean weight gains for children in the worst nutritional condition probably reflect the fact that those with clinical diagnosis received greater inputs," could be equally explained by the phenomenon of regression and sensitivity of weight-for-age discussed above.

What the above discussion suggests is the need to take into account the expected changes in a population. If the improvement is above or beyond what is explained by universal population characteristics, then we may begin to talk of impact. Development of "characteristic curves" which hypothesize relationships such as how nutritional status varies with age, and clarifying expectations on the spontaneous improvements of the malnourished and their higher sensitivity in terms of weight gain in response to feeding, is paramount.

Another problem with most evaluations reviewed was that they did not consider the social and economic factors, or physical phenomena which were taking place during the intervention as competing explanations for findings. The realization of the complexity and volatility of environments in which the programs are operating requires consideration be given to these historical and environmental factors. Unfortunately, when evaluators enter unfamiliar environments as outsiders and have little knowledge of what secular events transpired during the course of the intervention, it is difficult to account for such competitors. Furthermore, the lack of adequate controls in many of the studies further compounds this problem.

There are numerous other competing explanations and threats to internal and external validity not addressed above. The literature discusses these issues, which caution the reader to be skeptical of evaluation findings which stem from poorly conceived methodologies (see Sahn and Pestronk, 1981). These include attrition and addition effects, Hawthorne effects, multiple-intervention interference, selection biases of treatments or controls, as well as halo effects. Where it is possible to control for such factors, that is optimal. If not, the types of biases which may creep into an analytic procedure must be reconciled and discussed.

The perils of measuring nutritional impact as discussed above do not militate against striving toward that end. It does admonish against making program decisions based on hastily conceived methodologies which may grossly over- or underestimate improvement in nutritional status.

The task at hand is to minimize the operational pitfalls of doing nutritional impact studies. To begin with, it is paramount that individuals called upon to perform such tasks have a high level of expertise in the intricacies of determining nutritional impact. This requires a working knowledge of measuring nutritional status, evaluation methodologies, and data collection and analysis techniques. But more important, a sensitivity to and understanding of problems extant in doing field studies is paramount. This will enhance the possibility of overcoming the pitfalls of past efforts.

Even more important is the need to abandon the "fishing expeditions" in which a set of data is uncovered, followed by exploratory data analysis, which leads to developing an analytic procedure for examining those data; and thereafter the construction of a model to explain the findings. This approach is backwards. It has

appropriate application in some fields, but not so in nutrition evaluation.

B. Alternative Approaches to Measuring Nutritional Impact

The alternatives for future impact evaluations are twofold. The first is to undertake a series of impact studies, planned and executed to test predetermined hypotheses about program performances. The second is to set up built-in monitoring and data collection systems which would not only allow for continuous evaluation of program impacts, but also have major implications on program operations and the achievement of objectives. Adoption of this latter option is strongly advised.

Concerning the first strategy alluded to above, it is recommended that such field experimentation best be considered in the domain of "evaluative research" as defined earlier. These studies should be undertaken to answer specific questions which relate to program design or outcomes. Emphasis should be placed on exploring unknowns which have direct relevance to policy and decision making. Searching for impacts, and a determination of reasonable expectations of various types of programs, fall in the realm of this type of undertaking.

Needless to say, it is strongly recommended that any attempt to set up evaluative research to determine impact be longitudinal, as opposed to the often relied-on cross-sectional studies, that are almost without exception indeterminate. One should begin with an "experiment" in mind, and a way of testing hypothesized outcomes. Data are then collected prospectively according to the nature of the design. Analysis follows, using the tests prescribed by the predetermined evaluation design.

For example, it is possible to select a random sample of feeding centers, designate a matched control, and collect data at more than one point at a time. This type of cohort analysis is expected to yield results while controlling for outside variables as much as feasible. In fact, this approach would provide substantive findings within 12 to 18 months--not an unreasonable time frame as compared with efforts such as those in the Philippines, where months were required to organize data that were compiled hastily during four weeks in the field.

This approach is commended over the retrospective and cross-sectional designs reviewed. Reasons include that a prospective design allows one to keep track of new enrollees and drop-outs, which can have important implications for findings. Similarly, it is reasonable to collect data on program variables (e.g., how many times the child was fed, what was the size of the ration, what other services were provided) and keep track of environmental phenomena which may influence outcomes. Although some of this information may be available in retrospective studies where teams extract information from existing records compiled at feeding centers, experience indicates that it is usually not, or at least not accounted for by evaluations. It is also possible to institute quality control procedures in the collection of prospective data and make a premeditated choice of what indicators would most accurately reflect nutritional status changes in the population being examined.

Nevertheless, several drawbacks are obvious.

- (1) Outside evaluators would undoubtedly have to initiate, fund and operationalize such an effort, possibly biasing results. Hawthorne effects are likely to bias outcomes and thus represent one reason for advocating on-going monitoring/evaluation as discussed below.

- (2) If any anomalous event occurs during the evaluation period, which is very likely, the experimental design could be seriously compromised (e.g., epidemics, breakdown in food delivery, construction of a new road). The money spent would essentially be in vain. Similarly, unless a monitoring system was in place which could verify the delivery of services to the treatment group, and how that distinguished them from controls, it would be difficult to attribute observed changes to the program.
- (3) Little benefit would accrue to the project itself, at least in the short run, in terms of improving design or implementation. The learning potential to implementors and functionaries from the evaluation would be minimal.
- (4) But most important is that based on past reviews of nutrition impact studies, expectations of significant results should be modest. Detectable evidence of impact is difficult to find, given the imprecision of measures used and the numerous confounding variables to complicate the analysis. Conclusions about impact or the lack thereof would be hard to defend. And, undoubtedly, they would be subject to different interpretations.

These problems suggest the need for complementing such evaluative research. Validation of program impact is best based on continued and on-going measurement and analysis. This allows for results which may be solely suggestive or provocative based on a single measurement of change, to become objectively verifiable over time.

This concept leads to integrating the measurement of nutritional impact into on-going monitoring and data collection efforts. In this instance, evaluation serves two purposes: measurement of whether a program met its objectives, and provision

of information for decision alternatives and, hence, a course of action in the future aimed at improving program design and implementation. This is done by integrating the collection and interpretation of impact data into normal program procedures. A limited amount of impact data should be collected as part of normal operations. Routine analysis of the data will not only provide continual information as to whether a program is meeting its objectives, but can provide continuous feedback to functionaries, allowing them to adjust and alter their activities to facilitate the achievement of impact. This implies the need for iterative experimentation in the field.

Some very tangible benefits for and advantages to building evaluation into a process of continued data collection may be identified.

- (1) On-going analysis of impact data will allow for continued verification of program impacts, reducing the doubt that has plagued findings of large-scale evaluation studies.
- (2) It can be expected that programs which monitor themselves and thereafter have such information fed back on a regular basis, will be more conscientious in accurately collecting data and will be more motivated to assure the success of their efforts.
- (3) It will encourage the use of evaluation findings as a normal part of management and decision making functions, and will encourage alterations and innovations in program design.
- (4) Findings will be more readily accepted by functionaries and implementors, than if outsiders are viewed as making unfair and inaccurate assessment of the project.

Finally, the development of on-going monitoring/evaluation systems does not preclude summative or outside evaluation of program impacts. Quite to the contrary. With data systems in place, hopefully standardized in terms of variables such as measures employed, consultants or external evaluators may perform comparative analyses of program impacts, or do retrospective evaluations of program impacts over the years. In fact, if such self-evaluation systems were previously installed, it may well be that the concept of intermittently sending teams from Washington into the field, as in the Generic Scope of Work, would be totally appropriate.

Built-in monitoring/evaluation systems have two major negative attributes. The first is financial cost. There is a need for greater administrative overhead for the training, and possibly the addition, of staff to assure that the monitoring/evaluation system is operating as planned. Nevertheless, it must first be recognized that the voluntary agencies at present have developed monitoring systems to follow the flow of food. Although it would require an additional effort to collect, process, and analyze impact data on a regular basis, it would be viable given additional administrative capacity. In fact, it is arguable that at present the ratio of administrative overhead to food may be too low to allow the commodities to be programmed effectively. And when one considers the large sums of money being spent trying to evaluate nutritional impact, and largely to no avail, the expenditures on initiating monitoring/evaluation systems may prove cost-effective.

The second drawback is the contention that functionaries are not able to evaluate their own activities objectively. If the system is set up in a non-threatening way, and if those centers which register serious problems are not scorned, but rather worked with

in a positive fashion to improve performance, such difficulties may be overcome. Of course, this requires that middle-level managers be skilled and capable of and committed to interacting with village-level personnel.

The implementation of monitoring/evaluation systems must be evolutionary. They can only be installed in a limited number of areas at a time, due to the constraints on training both middle-level management and village-level workers. In addition, the system should be designed to collect a minimum of information. Over time, if a clear need is manifest, a more sophisticated system can be installed to respond to specific information needs for decision making.

The time for beginning the process of instituting built-in evaluation systems is right. The voluntary agencies are increasingly aware of this need, and concerned with providing cost-effective programming. In addition, the concern with using increasingly expensive food aid wisely is a sentiment strongly held in Washington. And the dismal failure of proving or disproving nutritional impact of supplemental feeding has created an awareness among professionals of the need for innovative approaches.

A first cut at conceptualizing and setting up such a system was started by Community Systems Foundation in India (Miller and Pyle, 1981). It will undoubtedly need refinements. Nevertheless, it is a beginning point for further discussion and experimentation. The process of implementing such a system must be characterized by mutual learning of AID/W, the Missions and the voluntary agencies. It is felt the long-term dividends will overshadow short-term apprehensions and discomforts.

V. FOOD-FOR-WORK PROJECTS

The procedures used for evaluating food-for-work projects have been similar in most of the evaluations reviewed. Described in its most simple fashion, evaluation teams select a number of sites for review on a random or ad hoc basis. They interview managers and beneficiaries, hear anecdotal and impressionistic reviews of project procedures and impacts, and carefully observe programs for the presence or absence of characteristics which facilitate or constrain success. Armed with that information, teams provide descriptive, informative, and insightful stories which recount their experiences. This leads to the formulation of generalizations about project operation. Recommendations for program improvement, expansion, or contraction are also provided.

This procedure seems entirely reasonable and appropriate given the fact that rarely are detailed records available concerning project inputs and outputs, let alone the achievement of purposes and goals. The decentralized nature of the program makes this sort of subjective judgment by evaluative teams all the more reasonable.

Despite the laudable job done by evaluators working under severe constraints, some conceptual and methodological issues demand attention in order to improve on the evaluation procedures relied on in the past. First, there appeared to be confusion as to what food-for-work programs are designed to achieve, and how achievement is to be measured. The result was that the indicators used to assess project impact at the purpose and goal level are often inappropriate. These problems will be discussed below under the heading of Project Impacts.

The second major hindrance to measuring the impact of food-for-work projects was that problems with project design and implementation were so overt and serious as to preclude considering impacts. These considerations fall under the rubric of program processes. Objective clarification and making project design reflect purposes and goals is one task related to evaluation which demands increased attention. Another distinct need is to formulate a monitoring/evaluation system to assure program processes from the input to output stage are transpiring as planned. Unlike the monitoring/evaluation system suggested for measuring nutritional impact at the purpose level, this system will focus on intermediate indicators which fall in the category of process variables measuring the transformation of inputs to outputs.

A. Process Evaluation

To date only a limited number of Title II projects with FFW have been the subject of the evaluations which have been originated by AID/W during the past two years. The purposes of the food-for-work (FFW) programs reviewed have been diverse. Their confluence, however, is manifest in terms of the common goal of fostering economic development. This is a reflection of the fact that the Food for Peace Handbook stresses " . . . the achievement of needed agricultural/economic and community improvements by providing commodities to support the labor of unemployed and underemployed local workers." Nutritional need is not given priority, nor is the concept of humanitarian relief.

As mentioned previously, a primary concern of process evaluation is to clarify program objectives. As in SF and MCH programs, evaluation teams made a determination of program purposes/goals on the basis of reviewing program plans and discussion with in-country personnel. In

some instances these were stated explicitly, while in other cases not. However, what became clear is that programs and evaluations often confused or did not distinguish between short-term consumption (welfare) goals of transferring income and long-term economic development goals. In other words, the creation of employment through food-for-work is geared to:

- (1) relieving short-run or seasonal shortages of food/income, and/or
- (2) reducing long-term structured unemployment and increasing economic growth.

These two aims are not necessarily antithetical, but often conflict when programming decisions are made. The fact that FFW projects have tried to serve both the welfare and economic objective function has resulted in confusion and impeded efficient resource allocation.

As such, the setting of program objectives (or more specifically, designating project purposes) should not represent a task relegated to a perfunctory chore by evaluators. The significance of dealing superficially with objectives is twofold: (1) it leads to ad hoc programming, or alternatively, conflicts in programming emerge due to the inherent incongruities in differing objectives; and (2) this in turn impedes efforts to evaluate programs.

In other words, the fact that objectives of FFW programs are often nebulous results in programming decisions being misdirected. But more to the point of this document, impact evaluation is derailed unless consensus on objectives to be measured is reached among implementors and evaluators.

As with all evaluations, once objectives are defined, the next step is to review the program plan to see whether prescribed activities are congruent with objectives. If so, it is reasonable to carry on

with the evaluation. If evaluators find that objectives are obscure and not reflected in coherent program plans and designs, they should focus their attention on such a problem, rather than continuing on in the evaluation process. Falling within the domain of evaluating program plans is to consider how well FFW is integrated with other developmental assistance. That is, evaluations should examine: if the selection of FFW projects is done in concert with government decision makers and agencies (see the Lesotho evaluation for a rare example of how this process takes place); and the degree to which FFW programs are complementing USAID priorities and strategies as prescribed in the CDSS; and more important, whether the activities in areas such as irrigation, flood protection, reforestation, and so on, are integrated with other USAID assistance being provided by the Offices of Agriculture and Rural Development. In answering such questions, evaluations will elucidate whether Title II food is a resource being used most propitiously. By piggy-backing FFW with other developmental assistance, FFW will alleviate its greatest constraints such as poorly conceived project designs, lack of technical expertise, and lack of equipment and raw materials to complete assets and infrastructure.

This leads the evaluator to examine the implementation process, once objectives and plans have been reviewed. Project inputs of proper quality and appropriate quantity must be provided for a food-for-work project to succeed. Thereafter, these inputs, including food, complementary inputs like tools, equipment, cash, and trained staff must be used efficiently and correctly in work activities to result in projected outputs. Such outputs may include roads, wells, hospitals, acres of cleared land, and so on. The outcome of producing outputs ranges from achieving purposes such as increasing employment, the ability to cultivate more land or the same land more intensively; improving access to markets; to just providing an income transfer to

the indigent. The higher goals are either primarily directed toward meeting humanitarian objectives, or alternatively are designed to promote economic growth and development.

Without exception, the evaluations identified breakdowns in the provision of inputs. Their supply in proper proportion is sacrosanct to FFW activities realizing designated outputs and thereafter purposes. Food supply breakdowns, lack of needed complementary inputs, poor on-site supervision, and low labor productivity, were among the problems identified. Similarly, the fact that projects often did not reach completion (i.e., achievement of outputs), or that poor materials, unsatisfactory construction techniques, or lack of maintenance impaired the value of assets created, were constant themes. Often, it was even impossible to determine how much food was provided to beneficiaries.

Given the seriousness of the management problems connected with FFW programs, the development of monitoring and management information systems for FFW is a prelude to impact evaluation. Sending evaluation teams to assess program processes (i.e., objectives, plans, operations), focusing on the degree to which inputs are transformed to outputs is strongly recommended. As a necessary adjunct to such an "evaluation" task, the provision of technical assistance and financial support for setting up basic management procedures for FFW projects is advisable.

It is recognized that efforts to improve the plan of operations and develop management information systems is costly. The need for doing so is overt in many instances. If movement in this direction is not considered viable by AID/W, it would seem fatuous to continue evaluations of FFW under the pretext of improving the quality of existing programs.

The outcome of the decision will reflect how serious AID is about treating food-for-work projects in specific, and food aid in general, as a development resource warranting as much attention as other capital development projects.

B. Impact Evaluation

Once objectives have been clarified, plans reviewed, and a monitoring system established to indicate that there is a reasonable expectation that programs are being implemented as planned, it is appropriate to measure program impacts at the purpose and/or goal level. Until these prerequisites are fulfilled, impact evaluation is premature.

Impacts, by definition, correspond to the stated objectives which must be reflected in project design. To illustrate, let's consider on the one hand the objective of meeting a crisis of acute food shortages or unemployment via a program designed to supplement or replace usual income sources. In this case, the program, which I will label Type A, aims to deal with acute poverty or crisis unemployment. Often this may be due to conditions such as flood or drought which may have destroyed the livelihood of a community, or perpetual food shortages due to a variety of socioeconomic determinants. This type of program is best considered a relief or welfare program. Concern focuses on getting food to the people. Little benefit accrues to those other than the destitute from such short-term redistributive efforts. The creation of lasting assets is given second or third priority. Planning for such relief activities centers on the ability to respond rapidly to crisis situations, or the identification of the most impoverished communities.

Meeting welfare objectives demands limited complimentary inputs and technical supervision. The goal is largely humanitarian. The

rural works projects serve to minimize the "handout" mentality, occupy idle labor, and in some instances help restore land which may have been ravaged. Nevertheless, attention is focused on short-term consumption goals, as all who are needy are invited to participate. The marginal product of labor is often near zero, and is of minor concern. Little is done to foster economic development or address persistent poverty or structural unemployment.

Whether food reached the poor and indigent during the period of food distribution is the appropriate impact measure of such a project. Measuring income effects in the short term indicates the value of the resource transfer. Counting the number of rations delivered and the value of the ration, are also important indicators which substantiate the achievement of welfare objectives. Such an approach corresponds, for example, to the CRS India FFW program and the CARE Philippines program until this year.

A second type of program may have as its primary objective maximizing economic returns, and stimulating economic growth. I will label this a Type B program. In this instance, Title II food is used as in-kind payment in rural works projects. Construction of assets which will have a high rate of return on investment is paramount. The main concern is with the stream of economic benefits which will accrue from the construction of a productive asset over time. Classical steps in project preparation should be followed. These include: (1) identifying investment opportunities; (2) doing a feasibility study to determine both the technical possibility and economic viability (this brings together engineers, technical specialists, financial analysts and economists); and (3) performing a project analysis. The latter step of analysis (i.e., cost/benefit) should be based on social rather than market prices, and should consider substitution possibilities between labor and capital, weighting values

heavily in favor of the former if long-term employment is considered a priority. But regardless, the benefits in terms of short-term employment and consumption are given secondary importance. Therefore, evaluation must focus on what transpires after the construction phase of the project. Instead, the concern is with the impacts during the operational phase, when the asset is designed to have economic multiplier effects, and food is no longer being distributed.

Determining the redistributive outcome of projects directed toward asset creation and enhancing long-term productivity and agricultural output is difficult. The issue is dependent upon who owns and reaps the benefit of the asset. If wells are dug in concert with land reform, or as part of a scheme where land-reclamation results in newly cultivatable land being provided to workers, then new assets will benefit the landless and destitute. Unfortunately, it is more likely that existing landowners will derive the greatest benefit from the creation of new assets. Income distribution will therefore not become more equitable. The contextual nature of the impacts on income distribution in addition to trade-offs with other objectives such as maximizing economic returns must be considered. Resolving potential conflicts between such benefits is best done on the basis of project appraisal techniques, attaching socially determined values (shadow prices) to such benefits.

Another distinguishing characteristic of programs designed primarily to increase agricultural output through the building of directly productive assets such as bunds, irrigation works, and land-clearing and reclamation, is that they require a good deal of advance planning and technical assistance. There is also a demand for complementary inputs in terms of materials and tools to complete such a project, as well as skilled labor and management, and supervisory personnel. In fact, it is recommended that such projects receive the same type of status and scrutiny of other capital

development projects. More often than not, the degree of material inputs and expertise required will be large enough to justify such an approach, and the need for concerted planning and coordination with overall development strategies is commensurate with any other aid projects.

Unlike a relief orientation (Type A), this Type B program will not necessarily serve communities or individuals in greatest economic need in the short or long term. The returns on an investment may be much greater in a region endowed with natural resources (e.g., fertile soil), or that already has a critical level of social and economic infrastructure (e.g., market access). This will allow investments to be synergistic. While relief projects may, in their quest to reach all the poor, employ so many individuals as to reduce the marginal productivity of labor to zero, this is not the case in a project aimed at maximizing economic returns, and therefore interested in labor productivity. Therefore, the degree of short-term employment created may be less, and projects may be less labor-intensive than those with primarily a welfare objective.

It is necessary to evaluate Type B projects not exclusively in terms of the number of workers employed or rations delivered during the construction phase. Rather, we are interested in multiplier and other economic effects which transpire during the operational phase of the asset or infrastructure created. This is the period when the provision of food as in-kind payment has ended. The variables of interest are mainly agricultural output, structural employment, and income distribution and price effects. While these impacts may prove elusive to quantify, proxy measures during a project's operational phase are suggested. For example, in Bangladesh the impacts of FFW were measured in terms of:

- (1) Production Effects - can increased agricultural output be attributed to the program?

- (2) Marketing Effects - has increased access to markets affected the types of food produced and/or market prices?
- (3) Employment Effects - has there been increased employment of newly-irrigated or cultivated lands?
- (4) Flood Protection - is there less crop loss due to flooding?
- (5) Land Value - have land values increased/decreased?
- (6) Consumption Effects - how have changes in income, prices, and food production affected consumption?
- (7) Distribution of Benefits - which groups of people have benefited from the above changes?

Additionally, we may want to ask whether projects have encouraged a sense of community self-reliance. Has an organizational structure emerged in villages which can generate new self-help initiatives, and at least maintain the assets created?

Answers to such questions are necessary to determine whether FFW projects have impact on economic and social development. Unfortunately, counting the number of jobs created does not meet this end. Even recording the miles of roads built may mean little if, for example, its secondary effects result in increased migration, and no positive marketing effects result; or, as is often the case, lack of maintenance on poorly-constructed roads results in their destruction after the first monsoon.

Despite the need to evaluate programs in terms of their impact on economic development, this represents an arduous task. It is not feasible to explore such impacts on all or most of the FFW projects. If this be the case, what to do?

It is recommended that AID/W in conjunction with the USAID Missions and voluntary agencies undertake a series of in-depth longitudinal research studies. Efforts should be made to distinguish groups of

projects with similar characteristics, and consider the range of impacts which are anticipated to accrue to a project. The emphasis should be on gaining insight into the benefits and constraints of different FFW programming strategies. Although it will be impossible to examine all gradations of project types or environments, representativeness should be sacrificed for solid research design which will minimize ambiguity of results. The cumulative knowledge of these studies will thereafter be used to guide future selection of FFW strategies to meet designated objectives, and enhance program design specifications.

It is only by evaluating the "operational phase" or "secondary effects" of a project that anything meaningful can be concluded about developmental impacts. As emphasis of food aid shifts toward facilitating national growth, and away from welfare objectives, this need becomes all the greater.

C. A Related Note

In addition to relief projects (Type A) and those designed to create long-term productive assets (Type B), other categories of food-for-work projects may be identified. Some are intended to provide the poor with new skills and often promote small-scale entrepreneurship. FFW is also often used to develop social infrastructure, such as school buildings, health clinics, and housing for the poor. Developing economic infrastructure, such as roads and marketing facilities, is also a popular use of FFW. Deciding which projects to sponsor must be seen as a function of objectives. Those objectives must first distinguish between short-term consumption gains and long-term production increases. Likewise, the tradeoffs between economic returns of food donated, the degree to which employment is created in the short term versus long term, and the effects on income distribution must be carefully considered.

A first cut at exploring those relationships is found in Table 4.

While this paper is not the place to discuss the intricacies of such trade-offs, it is paramount that they be clearly understood by decision makers in the field. This will reinforce the necessity for concisely delineating objectives of FFW programs, or the mix of projects to be undertaken. Recognition that there are conflicts, for example, between short-term welfare and consumption objectives (i.e., income transfers) versus long-term production (i.e., agricultural output) objectives is important. It will provide a conceptual basis for project design, be employed as criteria for approving or rejecting proposals, and allow evaluators to attempt to measure impact in terms of appropriate concerns.

To illustrate, the Philippine evaluation noted that the CARE program had been "primarily to assist disaster victims by providing food commodities that will generate work for reconstruction and rehabilitation." However, as of this year they have "redirected efforts toward developmental projects rather than disaster relief. Developmental projects are defined by CARE as those which help increase food production. Due to this change, there should be the expectation that the nature of projects undertaken will be different. In fact they may be less labor intensive, and rely on the employment of more able-bodied individuals; or be undertaken in relatively more prosperous communities. The justification is that the canals, roads, wells, etc., will increase agricultural output. Thus, once the project is completed the benefits will accrue to the whole community, including the poor. It is, therefore, not necessarily incongruent with program objectives that many of the workers receiving food owned radios and have electricity, as suggested by the evaluation team. Nor is the program a failure because "the income effects of the program are minimal." Given CARE's new orientation, the evaluators must look beyond the construction phase

TABLE 4

OBJECTIVES OF FOOD-FOR-WORK PROJECTS

<u>PROJECT CATEGORY</u>	<u>EMPLOYMENT CREATION (and Income Effects)</u>	<u>REDISTRIBUTIVE EFFECTS</u>	<u>ECONOMIC RETURNS (e.g., Agricultural Output)</u>
Relief Programs	High in Short-Term None in Long-Term	Short-Term in favor of the poor	Low
Productive Assets e.g., wells, cleared land	Low-Medium in Construction Phase High in Operational Phase	Construction Phase in favor of the poor Operational Phase dependent on Owner of Assets	High
Training and Small-Scale Manufacturing	Low in Short-Term High in Long-Term	In favor of the poor	Medium to High
Economic Infrastructure e.g., roads, market facilities	High in Construction Phase Low-Medium in Operational Phase	Short-Term in favor of the poor Long-Term variable	Medium
Social Infrastructure e.g., health clinics, community centers	Low-Medium in Construction Phase Low in Operational Phase	Dependent upon who uses facilities	Low

to when the assets are operational.

D. Nutritional Impacts

1. Construction (i.e., Feeding) Plans

It is feasible to measure the nutritional impact of projects designed primarily to meet short-term consumption objectives. The evaluation teams noted that improved nutrition is often listed among a multitude of purposes for FFW. For example, in Kenya nutritional benefits from food wages are identified as a goal. Mortality, morbidity, and weight-for-age of children are the indicators suggested for measuring nutritional effects. In such cases an argument can be advanced for the appropriateness of measuring the nutritional impact of such interventions.

The evaluation literature, in general, reports few examples of attempts to quantify nutritional impact of FFW schemes. Exceptions do exist. Note the WFP (1976) study which found that despite participating in FFW projects, families met only 50 to 60 percent of their caloric requirements.

Nevertheless, this paper will not enter into a discourse on the subject of methodologies appropriate to measuring the nutritional impact of food-for-work projects. I feel it unnecessary and inappropriate to attempt to determine the nutritional impact of most programs. This is not to suggest that empirical microeconomic research should not be performed to gain insight into consumption and nutritional effects of FFW, based on income, substitution, and price effects. Quite to the contrary.
Microeconomic analysis of household

expenditures and consumption of families provided with an FFW ration is needed. And it may even be reasonable to gather anthropometric information as an adjunct to such an effort. Nevertheless, it seems premature and poorly conceived to expect this to be done on a country-by-country basis. Rather limited evaluative research efforts aimed at understanding questions about household dynamics, or the values of in-kind versus cash payment for work are needed. The reasons are as follows:

- a. FFW projects should focus attention on creation of assets and infrastructure which provide the means and hope for economic growth and improved income distribution. Since most projects are of short duration, seasonal in character or reach only a fraction of those in need, it is imperative that programs have multiplier and long-term economic impacts. Otherwise, some communities will necessitate continued feeding. The obvious problems of creating dependencies and suppressing private initiatives in such welfare schemes is only one problem; the other being that we (the donor countries) cannot feed the world's indigent. Thus, FFW projects must look beyond short-term consumption effects during the construction phase.
- b. The ration in an FFW project is designed to remunerate the worker for his labor. Whether or not the payment is entirely in food, or partly in cash (which is often the case), the food ration is best considered as any other type of income transfer. In order to adjust for the potential undesirable effects of in-kind payment, two practices are noted which compound the difficulty of measuring nutritional impacts: families re-sell or barter food, and they also readjust food expenditures given the availability of the ration.

- c. The marginal propensity to consume goods other than food (e.g., fuel, clothes, shelter, education) may be much more significant in toto than increased food expenditures. Looking at nutrition, as opposed to the entirety of household purchases would provide misleading information on microeconomic and short-term consumption effects.
- d. Food-for-work projects often reach families or individuals for short periods of time. That is, participation is rarely constant (daily) over a long time-frame, and varies greatly from individual to individual. This makes evaluation very difficult. It would be necessary to examine differential impact in terms of varying degrees of participation.
- e. FFW rations are targeted to the entire family unit. In order to assess directly nutritional impact of FFW, it would be necessary to measure nutritional status of all family members. This clearly represents a cumbersome and expensive effort.

2. Operational Phase

To my knowledge, no successful attempt has been made to measure the long-term nutritional impact of food-for-work. This is logical. To do so would require on-going surveillance of nutritional status after projects are completed and the intervention phase over. While I would continue to argue that measures other than improved nutrition should be accorded the attention of evaluations, it should be possible to design a research effort in which nutritional status of two matched villages were compared over time. The experimental village would be that in which a new road was built, new wells dug or the like; via FFW the control would be where there was no intervention. Observing nutritional

changes in the two villages, if a surveillance system was in place, would be an interesting endeavor. It would provide insight into whether the long-term developmental improvements brought about by FFW had any effects in nutrition. In addition, it would indicate whether the benefits do not reach vulnerable groups (an often-heard contention) or do not address an active constraint to the attainment of adequate nutriture (e.g., parasitic infections).

If AID/W decides to undertake such studies, it must be realized that such evaluative research requires a great deal of sophistication. I would also contend that such research must be subsequent to a commitment to improve the planning and implementation aspects of most programs around the world.

VI. ECONOMIC IMPACTS

In the unrelenting debate over food aid, proponents and opponents inevitably point to the economic impacts of programs to bolster their viewpoint. Traditionally, the major area of contention has been in the realm of program aid (i.e., Title I), as opposed to project aid where food is distributed through differentiated market channels (i.e., Title II). Nevertheless, as food aid has come under closer scrutiny by critics, and concurrently, advocates search further to understand the implications of their actions, economic impacts are becoming the domain of Title II programs as well as concessional sales.

This section of the report discusses those economic impacts which were addressed in the evaluations reviewed. The intent is to deliberate whether the topics covered by the evaluation were appropriate. Simultaneously, there is an attempt to identify the salient areas of concern, and which if any, should be the focus of further inquiry. The intent of the brief discussion that follows is to guide AID/W in deciding on what questions they should be asking. The need for more careful and sophisticated methodologies to evaluate economic impacts is also stressed.

A. Microeconomic Effects

When examining the potential microeconomic impacts of a feeding program at the household level, the primary variables of concern are how changes in prices and income alter expenditures and consumption. In the context of evaluation, the use of microeconomic data is two-fold:

- (1) to impute nutritional changes which result from a program; and
- (2) to gain insight into the dynamics of household economics, so

as to better design programs and related educational efforts based on empirical estimations of consumer behavior.

As for imputing changes in nutritional status, it is recognized that food expenditures are mediated by prices and income. In concert they determine the amount of resources the family devotes to their dietary objectives, and the extent to which that allotment can satisfy a family's nutritional needs.

1. Price Effect

Price effects on food consumption can be considered a function of (a) substitution effects and (b) income effects. The substitution effects are changes in purchases when consumers adjust their commodity bundle as a result of the relative price shifts of goods. The income effect refers to the changes in real purchasing power which occur when prices of a commodity in a consumer's bundle change. The most accurate method of determining such relationships is in terms of the Slutsky Equation. Its simplest form is as an elasticity equation which is summarized as follows:

$$e_{ij} = \epsilon_{ij} - E_i \alpha_j, \text{ where}$$

e_{ij} = overall demand elasticity for commodity i when the price of commodity j changes,

ϵ_{ij} = the pure substitution elasticity for commodity i when the price of commodity j changes,

E_i = the income elasticity for commodity i, and

α_j = the budget share of commodity j in the consumer's total expenditures on all commodities.

This relationship is discussed by Timmer (1981) and others in the economics literature on consumer theory. Note that E_i and α_j (income elasticity and budget share) are a function of income

class. The poorest people often devote 75 percent of their budget to food and have income elasticities for food around 0.8. This is all to say that the low income consumer is most sensitive to price shifts.

This report will not detail any further the relationship between changes in prices and food consumption of the poor. However, the above equation is presented to illustrate the theoretical basis for, and types of, empirical research which must be undertaken to estimate how price changes affect household expenditure patterns. Thereafter such price effects can be used to impute changes in consumption and nutritional status of individual household members. This final process of measuring intrafamilial distribution of foodstuffs falls in the domain of the nutritional anthropologist. But let it suffice to say, it is a precarious task, in and of itself.

MCH or SF programs will have little or no measurable impact on relative prices. Therefore, price effects are not of great significance. This was apparently recognized by the evaluations done to date which pay little or no attention to price effects. However, FFW may have direct and indirect effects on village level and regional food prices, which will determine consumer choices. Quantifying such empirical findings is difficult, and should remain outside the purview of Title II evaluations. Rather, it is the subject for research economists. However, an understanding of the principles of consumer theory can be applied to the formulation of program design. It is not recommended that the evaluation of price shifts and their impact on the consumer be a subject of short-term excursions out to the field.

2. Income Effects

In the above discussion it was illustrated how price changes effect consumption via substitution and income effects. There are also pure income effects in which we are interested in how changes in income (in-kind or cash) alter the commodity bundle purchased. The model of how consumers change their purchases as a result of increased income is driven by a number of variables. For example, expenditure patterns are a function of individual preferences, which stem from endogenous factors such as wealth, form of income (e.g., in-kind), period in which income is received, household member who receives income, cultural habits, and so on.

However, unlike price effects, all the evaluations reviewed included a discussion of income effects. This suggests that evaluators felt competent, and that it was legitimate, to consider program impacts partially in terms of the income transfer. This leads to asking two questions:

- (1) were the methodologies employed to measure income effects adequate? and
- (2) is this a fruitful area of inquiry in terms of MCH, SF, and FFW programs?

Concerning the first question, the answer is a qualified no. Firstly, measuring income, especially in developing countries, is a precarious task, to say nothing of increments in the form of in-kind payment. Numerous problems are obvious, even if one assumes that it is possible to quantify accurately the amount of food received by each family as a result of program participation. Does one determine the income effects on the basis of the market price of the same commodities such as soy-fortified bulgur, or on the basis of local equivalents? Does one attempt to cal-

culate a shadow price for these donated commodities? The alternative possibility of calculating the income effect on the basis of the value of local commodities which supply similar caloric values, and are not purchased because of the in-kind transfer, must be considered. It is obvious that there is a need for the development and consistent application of a methodology for valuing in-kind payment. This is especially so if such an endeavor is going to be the subject of future evaluations and a comparative analysis of findings.

Concerning the second question posited above, once again the answer is a qualified no. The reasons are manifold. But most fundamental is that despite vociferous assertions otherwise, I would counsel against trying to justify feeding programs on the basis of an income transfer. Simply put, in-kind transfer with donated food is cost-ineffective in more instances than not. The price of local equivalents or substitutes is usually less than the value of Title II foods when one considers procurement, transport, and overhead costs. Exceptions to this generalization often include milk and vegetable oil.

To illustrate, if the cost of sending one bag of wheat to Bangladesh is \$X, and the same \$X can buy 1.5 bags of wheat at local markets, food aid is not an expeditious way of transferring income, and the justification for doing so is precarious.

Another oft-heard reason for measuring income effects is that improved nutritional status is mediated not only through the provision of the supplement, but by the income increment. This is undoubtedly the case. But why use an indirect measure of income to impute nutritional changes if more direct measures of nutritional status are available and more facile?

Despite the above deterrents to examining income effects, there are

reasons for looking at how they affect household expenditure and consumption patterns. For example, we might rely on microeconomic analysis in considering to what extent the food provided in various types of feeding programs (e.g., with or without education) is additional to or substitutional for family food expenditures. Another possibility is to use microeconomic data to determine whether a tied subsidy is "more beneficial" than a cash transfer, as measured by its effect on individual and/or household food consumption; and what consumption changes other than increased food expenditures are manifest by the family. These types of questions must be empirically determined. In the case of comparing in-kind versus cash payment, one could initially use multiple regression to compute income elasticities of demand for calories out of a tied versus cash subsidy.* Thereafter, multiply that factor by the proportion of household expenditure being spent on food. This would provide the analyst with the marginal propensities to consume (MPC) out of in-kind versus cash income. Such a calculation would have to be disaggregated by some income or wealth parameter. If the MPC out of in-kind income is greater than cash, there is clear justification for continuing food aid to reach nutritional purposes.

The use of household or expenditure surveys is appropriate for providing the types of empirical data identified above. Obviously, the cost and sophistication required for such endeavors militate against widespread application. Rather, specific research regimens should be formulated, limited to answering specific hypotheses and addressing fundamental assumptions which provide the conceptual basis for programming. This is contrasted with using such analytic techniques to assess success of a large-scale on-going program, as has been

* It is important to distinguish between income elasticity of demand for food and income elasticity of demand for food energy (calories). The difference between food expenditure and food energy elasticities is attributable to an increase cost per calorie.

the case previously.

B. National Economic Aggregates

Evaluations of P.L. 480 Title II have given little attention to the macroeconomic impacts. This is to be applauded. The major macroeconomic variables, national income, savings and investment, and inflation and unemployment are rarely affected in any major way by Title II. The reasons are manifold, but most fundamentally, the size of the transfer and magnitude of the resource are too small to have any measurable national macroeconomic implications. This is not to say that program food aid, especially Title I, does not have macroeconomic impacts. But non-~~fungible~~^{fungible} project aid will rarely exhibit any effect on economic aggregates.

The Ghana evaluation was one exception to the general reluctance to address macroeconomic effects at the national level. It is felt that the amount of time and effort put into preparing these sections of the report was generally inappropriate for a Title II impact evaluation. And in those anomalous situations where this information is considered vital for programming decisions, it should be collected as part of an effort distinct from the Agency's impact evaluation plans. On a more specific level, I would take issue when the Ghana evaluation states: "Title II loomed relatively large in the economy," when my calculations based on figures from the World Development Report, 1980 (World Bank, 1980) indicate that it comprised less than one tenth of one percent of the gross domestic product. The evaluation team also argued that since Title II substantially increased food availability, it was an anti-inflationary measure on a national level. Numerous questionable assumptions underlie this assertion, which would demand rather lengthy discussion. This is not an appropriate forum for doing so. However, let it suffice that in those rare circumstances where national macroeconomic aggregates may be affected by

Title II programs, that more considered empirical analysis and greater discretion is advised before making any unqualified judgments about impacts.

C. Inflation

Food aid has been referred to as a hedge against food price and general inflation. Specifically, the argument is that poor households have a high income elasticity of demand for food. If the income transfers represented by SF and MCH programs, or the remuneration of FFW projects was in cash, rather than in-kind food payment, the large increase in demand for food and other goods would result in inflation. The above scenario is based on the assumption that a food supply bottleneck would occur because food supplies are inelastic in the short term and additional income would not result in additional savings. Thus, the use of in-kind payment would all but eliminate these potential inflationary pressures.

In general, it is misdirected to pay attention to the role of SF and MCH programs in relation to inflation. The facts that the food donation is relatively small in comparison to the demand of a community as a whole, and that this food is distributed through differential channels, makes discussion of such impacts unfruitful. However, there are persuasive arguments that in rural works projects, at least partial payment in-kind is an appropriate anti-inflationary measure. This is due to the otherwise large infusion of cash which would result from initiating a rural works project.

But rather than looking at potential impacts of FFW in the context of circumventing the inflation which would result from a monetary payment, it is recommended that such considerations be integrated into the planning process, and selection of sites in which in-kind payment (hopefully in conjunction with cash) is most warranted. Such a

decision would essentially be based on an estimation of elasticity of food supply in response to the increased demand brought about by the rural works project. Simultaneously, this would help guide the appropriate mixture between food and money supplied to workers.

D. Production Disincentives

The most significant allegation leveled at food aid has been its potential disincentives on agricultural production. Food aid is alleged to lower prices for commodities and reduce incentives for local production. Evidence generally indicates that food aid which is programmed through differentiated market channels such as schools and health clinics has little impact on food prices. Of course, the amount of food aid in comparison with total production, the price elasticity of demand (how much additional demand may be created if prices are lowered), the price elasticity of supply (how farmers' production decision will be affected by price changes), are the determinants of possible disincentive effects. But given that most Third World governments usually drive a wedge between prices to producers, prices to consumers, and international prices, it is unlikely that project food aid will have significant impacts on production. Furthermore, if food aid is in place of, rather than additional to, "commercial imports," once again we may expect little potential for disincentives.

Two evaluations addressed the disincentives issue at the national level: those from Ghana and Lesotho. Both did so in a qualitative way, making judgements based on the magnitude of the food-aid contribution in comparison with the local food supply. However, it was beyond the scope of such assessments to come to any definitive findings. Despite that the Lesotho evaluation was provocative and well reasoned, it is not advised that future Title II evaluations examine such factors on a country-by-country basis. The reasons mentioned in the above

paragraph, coupled with the relatively small size of the transfer in comparison with most countries' production and/or commercial and concessional imports would make most of these pursuits futile.

However, regional disincentive effects mediated through lower prices attributable to large infusions of food into a community due to an FFW project is an interesting area for further inquiry. (This problem can be dismissed for SF and MCH programs.) It is also worth considering two mechanisms, other than prices, by which agricultural production may be suppressed. The first is centered around the decision of the marginal farmer to join in on a FFW brigade rather than farm the land. If wages are high enough, it may draw farmers off the land in lieu of what are perceived as less risky and more lucrative alternatives.

A second possible disincentive not related to prices involves the formulation of policy at the national level. If a nation feels that food aid is a reliable source of development assistance which can forestall shortfalls in the near future, they may select policies which have a bias against agriculture. Instead of expending resources in a quest to grow food, the government may choose to invest elsewhere. This is not necessarily a bad decision. Intelligent investments may allow countries to develop productive potential capable of earning foreign exchange in which to purchase needed commodities. Unfortunately, this is often not the case, as foreign exchange remains an active constraint, and/or is not used to purchase the basic foodstuffs which will feed the population.

As with the discussion of inflation, it would not be worthwhile to examine the question of disincentives on anything but a select basis. The intent of such research would be to identify common characteristics of communities in which the infusion of food aid is expected to act as a disincentive. Simultaneously, efforts should be made to confirm

that such deleterious effects are not the rule, but rather the exception. But most important, concern about disincentives should be explicitly addressed during project planning stages, when considering issues like where to initiate FFW and level of remuneration.

E. Economic Development

Economic development has been described in a variety of ways. Among them have been an increase in national product or an increase in national product accompanied by an increasingly equitable distribution of wealth. Other indices like the Physical Quality of Life Indicator (Morris, 1979) which is a composite measure of literacy, infant mortality, and life expectancy is suggested as an appropriate measure of economic development.

However, I would argue that PQLI is more accurately portrayed as an indicator of social development. Given that in the 1970s the United Arab Emirates had a PQLI of 34, ranking 112 out of 150, and was simultaneously the richest country in the world in terms of per capita income (\$15,368), it is obvious that there exists divergent viewpoints in measuring a country's state of development. Nevertheless, I will propose three parameters to consider whether a program contributes to economic development. They are whether:

1. it causes an increase in Gross Domestic Product;
2. it does not skew further the distribution of wealth and income; and
3. there was some transfer of knowledge or technology which will result in improvements in the economic or social fabric being self-sustaining, or, in other words, development involves helping people help themselves.

While this is not the forum to expand further on such issues, those interested in Title II feeding programs often list economic development

as a primary goal. This implies that evaluations must consider the success or failure of such efforts in terms of that criterion. However, this term is rarely defined, nor are the linkages between achievement of more modest objectives (e.g., improved nutritional status) and economic development stated.

To amplify, it can be argued that MCH and SF programs which are designed to improve nutritional status may lead to economic development. The argument is simply that if the program achieves its purpose of better health and nutrition, it will result in an improved quality of human resources in terms of physical and mental capacity. In turn, this will lead to a more productive and prosperous society. Such a theory is impossible to verify in quantitative terms. But more importantly, it should be recognized that there are a number of underlying assumptions, which make the linkage between improved health and economic development tenuous. Therefore, to justify programs or dwell on the potential impacts in terms of economic development will prove unproductive. Rather, I would suggest that we stress the contribution of health and nutrition interventions to human welfare, while concentrating on imparting educational messages on food and social practices which will have lasting impact. But it seems preposterous to expect MCH feeding programs to address any of the structural impediments or active constraints to the achievement of economic growth with equity. Acting under such false pretenses can only prove counter-productive.

School feeding programs with educational objectives, on the other hand, are legitimately justified in terms of their potential economic impacts. If programs can be shown to increase education, the linkages to economic development are well substantiated. Specifically, there is a convincing literature on the economic returns to primary education. It is therefore reasonable to make a qualitative assessment that economic development goals are being furthered, if evidence can be amassed that educational objectives are being met.

Finally, the linkages between the achievement of FFW purposes (in terms of the logframe), and economic development are most apparent. The one contestable issue in this context concerns the possibility that FFW projects will create further structural inequities. While many developmental economists will point to the inevitable patterns of the poor becoming more impoverished during the initial stages of development, before society's riches become more evenly distributed, this obviously represents an empirical observation worth trying to defy.

VII. EDUCATIONAL IMPACT

A number of the evaluations reviewed stated that the purpose and goal of a variety of feeding programs were to improve education. Take, for example, the school feeding program in Egypt, where increased learning and school performance were the goals, with increased enrollment, reduced dropouts, reduced absenteeism, and increased attentiveness as sub-goals. Similarly, the school feeding program in India has adopted education as its primary goal. On the other hand, other school feeding programs do not prioritize educational goals, as witnessed in Upper Volta.

In this author's previous review of the general literature, it was found that there is a paucity of evaluations which attempt to quantify explicitly the educational impact of the programs. Those that set out to do so have been plagued by methodological problems, and their findings are often inconclusive (Shortledge, 1980; Maxwell and Singer, 1979). Of the recent evaluations of Title II school feeding programs reviewed, few explicitly stated educational objectives as the primary focus. As a result, impact data in terms of improving education are a scarce commodity. Therefore, little guidance can be gleaned from experiences to date.

It is noteworthy that the only Title II evaluation which had quantitative data about educational impacts corroborates the difficulty of the task. To amplify, in Egypt, a qualitative assessment led the team to believe that the program was reaching its goals. But in reviewing a quantitative assessment performed by the Ministry of Education, it was noted that there were no conclusive data concerning the impact of the program on examination performance, attendance, or health status. They remark that this was due to inappropriate research design and data collection techniques.

Given the dearth of experience in evaluating school feeding programs in terms of educational objectives (as contrasted with the numerous nutritional impact evaluations which have been attempted), this area represents fertile ground for future attention. The most sensitive and precise measures of meeting educational objectives are: increased attentiveness and learning, improved school performance, and academic attainment by a greater number of children than if there were not a morning or mid-day meal. Documentation of such a finding would lend unrefutable proof to the value of school feeding. However, these indicators are difficult to measure. Therefore, it is recommended that AID/W initiate a limited number of evaluative research studies around the world aimed at providing information on such variables. Consultation with experts is suggested to design the most feasible research designs. The aim is to substantiate the hypotheses which link outputs to purposes.

Simultaneously, the SF program with educational objectives should consider collecting data on enrollment, dropouts, and absenteeism in the context of an on-going monitoring/evaluation system which covers all schools. These indicators are suitable proxies for measuring impact, are easily collected, and will provide satisfactory evidence of whether or not SF acts as an incentive to draw students to the classroom. It will also be worthwhile to set up some experiments where the incentive effect is compared between feeding and non-feeding schools. This should be done on a representative basis, examining differing environments and operational contexts.

The previous discussion has centered around the examination of the linkages between outputs (feeding children at school) and purposes (improved education). There is one more link in the hierarchical progression of program events: that of improved education leading to the goal of economic development. The literature is persuasive in this regard. Hence, it is not advised that AID pursue this avenue

of investigation.

Finally, one last comment is demanded in terms of educational objectives of school feeding. Simply, I would strongly recommend dropping the nutritional objectives of SF in lieu of educational objectives. The reasons are summarized as follows:

1. It will be difficult to measure nutritional impact of school feeding programs. It will be even more unlikely that impacts will be as encouraging as MCH nutrition projects. Justifying programs on the basis of nutritional impact is precarious and will probably result in the elimination of the program.
2. If social and economic development are priorities or higher program goals, nutrition objectives should be dropped and education objectives embraced. The reason is that the linkages between improved education and development are more evident than between improved nutritional status and development.

VIII. NUTRITION EDUCATION AND OTHER IMPACTS

A number of evaluation studies give a great deal of attention to the nutrition education, health care, sanitation and/or other components of a feeding program. Many professionals maintain the conviction that an MCH program must be complemented with outputs other than food. The conceptual basis for integrating programs is that services act synergistically to facilitate achievement of nutrition impact.

This issue of multi-faceted programs suggests two concerns. The first is whether there is solid evidence which indicates that a multiplicity of components do in fact act synergistically. Is the impact of the whole greater than that of the individual services being provided? Despite the acceptance of the need for integration, there are serious questions as to the validity of this dogma (see Sahn, 1980). But more to the point of this paper is the dilemma of how to evaluate the impact of such endeavors. For the sake of simplicity, this question is addressed below in the context of a supplementary feeding program with a nutrition education component.

I would argue that in an MCH program with a nutrition education component, the objectives remain the improvement of nutritional status. Therefore, impact evaluation demands the collection and analysis of the same types of data as a program without nutrition education or other complimentary outputs.

Next, the question emerges whether we should be concerned about the effectiveness of a program in altering attitudes and behavior or increasing knowledge. Do such areas of inquiry fall within the purview of impact evaluation? To answer these questions we must step back a moment and reconsider what we are trying to learn from the evaluation exercise.

One argument for measuring the changes in knowledge, attitudes, or behavior due to an educational component is that it serves as an indicator of the nutritional well-being of women, infants, and children. While there are linkages between, say, knowledge and nutrition, too many assumptions must be verified before considering any causal or direct relationship. For instance, one must show that knowledge translates into altered behavior. There must also be evidence that the altered behavior addresses what is truly an impediment to attaining adequate nutrition, and, that overcoming such a constraint is in and of itself sufficient to improve nutrition in the light of other more overwhelming and significant variables. Thus, I suggest that educational outcomes (e.g., knowledge) are too distal a measure to indicate anything about nutritional impact.

A second reason for measuring the effects of a nutrition education component is to determine whether, in concert with feeding, it improved nutritional status. This gets back to the question broached above. Do feeding and education act synergistically? This is an important issue, to which we only have partial answers. However, the search for such insights is best viewed as evaluate research. The quest is to ascertain the value of combining program outputs (e.g., feeding and education) in facilitating the realization of nutrition impact. If through such research we are convinced of the value of an integrated approach, our task is to concentrate on the planning and implementation of programs which follow such guidance. It will thereafter be sufficient to evaluate the impact of such programs in terms of nutritional status.

The point made above gets to the crux of the matter. Education is an output of a program, just as is feeding. The determination of a change in knowledge cannot be a proxy measure for nutritional impact, just as knowing that a supplement has been delivered to an individual is not an adequate indicator of improved nutrition. Rather, the number

of classes held as well as the degree of change in knowledge are intermediate indicators, although the latter is more proximal to improved nutrition. It must be recognized that a continuum of process measures do exist; nevertheless, those closest to suggesting that consumption or behavior was altered must not be mistaken for impact measures.

The preceding discussion refers not only to educational efforts in MCH and SF programs, but activities like home or school gardens, sanitation, health, and so on. To illustrate further, a causal chain of events, which in the case of gardens includes the production of food, is designed to increase food consumption and assist in reaching the objective of improved nutritional status. It is only logical that we measure impact at that level. Of equal importance is qualifying and quantifying those program attributes (processes) that contribute toward this end.

It is worthwhile determining whether a garden has been planted at a school, and whether the children consume some of that food. We also would be well advised to find out if a village has witnessed a proliferation of home gardens as a result of a nutrition education program, or whether the latrines which dot that map of the many villages are being used and cared for properly as a result of education classes. All of these process questions are fundamental to making sense out of impact findings, and will assist in the understanding of what are the ingredients to success.

It is generally advised that qualitative assessments of such program components are necessary in any impact evaluation. But despite the need to document how many classes were held, whether gardens are planted and so on, it is cautioned that such data could overburden a monitoring/evaluation system, if adopted as suggested. It is therefore advised that when questions exist about value of complementary inputs,

evaluative research projects be initiated on a limited basis. Subsequently, there is a need to: assure the integration of various components into program plans, and concentrate on being certain the implementation process is in accordance with planned specifications. That is a management responsibility. Evaluators may legitimately examine management techniques. But more important, if a monitoring/evaluation system exists, it will allow management to concentrate on improving the implementation process through feedback about program impacts. Simultaneously, such a system will assist evaluators in identifying individual projects which are distinguished by their successes or failures, allowing for further study to discern why.

A final point is germane to this discussion. A series of tangential effects or objectives of Title II programs are often highlighted by evaluators as criteria for measuring success. These include assisting people in learning to help themselves, developing a sense of community self-reliance, providing a mechanism for socialization of children and mothers, and so forth. It is difficult to measure such abstractions, and it is therefore precarious to justify programs in terms of these impacts. But more important, I would argue that factors such as community participation and initiatives are necessary conditions for achieving more tangible impacts, such as increased employment due to an FFW irrigation project or improved educational achievement due to school feeding. Therefore, it does not seem to be a propitious use of resources to evaluate these fundamentals. Rather, assurance should be provided that they are built into the project and are central philosophical tenets.

APPENDIX A

- Stewart Blumenfeld et al. (1981). The Impact of P.L. 480 Title II in the Philippines, 1970-1980, Agency for International Development, Washington, D.C.
- Judy C. Bryson (1981). An Evaluation of Food Aid Programs in Lesotho with Emphasis on the CRS Program, Agency for International Development, Washington, D.C.
- Judith W. Gilmore et al. (1980). Morocco: Food Aid and Nutrition Education, AID Project Impact Evaluation Report No. 9, Agency for International Development, Washington, D.C.
- Earl Jones, Sarah Berkowitz, and Robert Roussel (1981). Ghana Food for Peace (Title II) Evaluation, Draft, Development Associates, Inc., Washington, D.C.
- Roy I. Miller and William D. Drake (1981). P.L. 480 Title II Evaluation--Phase I: Assessment of Data Quality and Availability, Community Systems Foundation, Ann Arbor, Michigan.
- David Nelson, David E. Sahn, and Beatrice L. Rogers (1979). An Evaluation Report of the P.L. 480 Title II Program in India, Community Systems Foundation, Ann Arbor, Michigan.
- Richard Rhoda and Sandra Collier (1981). USAID/Cairo P.L. 480 Title II Program Review, Agency for International Development, Washington, D.C.
- (1980). Kenya Food for Peace Title II Evaluation, Draft Report, Practical Concepts Incorporated, Washington, D.C.

(1980). An Evaluation Plan for the P.L. 480 Title II Food for Work Program in India, Practical Concepts Incorporated, Washington, D.C.

(1980). Title II Food Aid Program Evaluation Generic Scope of Work for Country-Specific Evaluations, Draft, Agency for International Development, Washington, D.C.

(1981). Upper Volta Food for Peace/Title II Evaluation Draft Report, International Science and Technology Institute, Inc., Washington, D.C.

APPENDIX B

In 1980, as part of its own program of evaluating outcome, CRS headquarters collected data on some 28,000 children participating in its IMCH program in 1979 and 1980. The data set included participants from every region of the country. The data were compressed into 4 nutrition status categories--severe, moderate, mild malnourishment, and normal--and were hand-tabulated. The results showed that 58 percent had improved their nutrition status, 24 percent were unchanged, and 18 percent regressed.

From CRS's random sample of 28,000 a systematic sample of 5 percent was drawn for us by our contractor and the data were copied. Since the data were being computer-analyzed, the finer, more sensitive 10-level nutrition status scale was recorded rather than the less sensitive 4-level compressed scale. From this subsample was drawn yet another systematic sample of 238 cases for analysis in depth. To show the effectiveness of the subsampling procedure, it may be noted that in its narrative of the results of its analysis of these data the CRS staff comment that 25 percent of their sample are infants and that 13 percent are in the normal and mildly malnourished categories; in our subset, the respective figures are 24.4 percent and 13.9 percent.

71.8 percent of participants are under age 3 years, a figure which indicates that CRS is quite appropriately concentrating its effort on the most vulnerable age group. However, while targeting criteria call for inclusion only of children who are either severely or moderately malnourished, 18 percent of those 12 months or older were in the mild category. Fifty-nine percent of the 6 to 11 month olds are normal or mildly malnourished, but this is permitted in the plan for malnutrition prevention. The mis-targeting is most pronounced in the 1- and 2-year olds, 21 percent in each group. However, overall, those appropriately targeted constitute 68 percent of the sample. In terms of length of participation in the program, the sample, is not an unbiased representation of CRS's beneficiary population at the time because one of the criteria of selection was participation for at least 3 months. At the time of sampling, of the remaining group--the great majority--20 percent had participated 3-5 months, 39 percent 6 to 11 months,

and 41 percent 12 months or longer. Although 18 months is supposed to be the maximum duration of participation, 5 percent managed to stay in a little longer, usually 1 or 2 or 3 months, but as much as 28 months for one child.

In this sample, 125 of the 238 cases (53 percent) showed some improvement in nutrition status, while 24 percent were unchanged, and 23 percent regressed. For those who improved, the average gain was 2.1 nutrition levels in an average length of participation of the 11.7 months, or an average gain of about 0.2 levels per month of participation. The decline in nutrition status was much more pronounced in the 6 to 11 month groups than in the balance of the children: 59 percent of the former lost ground, while only 12 percent of the latter did. This had been noted also by the CRS staff in their analysis.

Analysis shows that those who enter the program in the worst condition have the greatest tendency toward improvement. Table 2 shows the cross-tabulation for all cases other than those entering with normal status.

Table 2

Gain, Loss, and No Change in Status by Entry Nutrition Level
(6-59 months)

Entry Level	Change			N
	+	0	-	
2-3-4	25%	18%	57%	60
5-6-7	64%	24%	12%	147
8-9-10	64%	36%	0%	25
				<u>232</u>

Given that, as noted above, decline in nutrition status has substantial correlation with age, it is necessary to determine if the result shown in Table 2 would be considerably different if the effect of the youngest (and most poorly fairing) group were removed. Table 3 shows the result of that analysis.

Table 3

Gain, Loss, and No Change in Status by Entry Nutrition Status
(12-59 months)

Entry Level	Change			N
	+	0	-	
2-3-4	41%	25%	34%	32
5-6-7	68%	23%	9%	128
8-9-10	65%	35%	0%	20
				<u>180</u>

It may seem that even with the effect of the 6-11 month olds removed, proportionately more of the moderately (levels 5-6-7) and severely (levels 8-9-10) malnourished improve and fewer decline.

The CRS data permit us to examine relationships between the nutritional outcome of the supplementary feeding program, measured as degree of change in nutrition status (coded DNL) and other measured variables, namely, age of entry into the program (ENTAGE), nutrition level at entry (ENTNL), length of participation (LOP), and nutrition level at the end of the period of participation (FINNL). The latter variable, however, is not necessarily the same as nutrition level upon "graduation" from the program because many of the children were still active in it; recall that 59 percent had been in for less than 1 year. Table 4 shows the simple correlation matrix (zero order) for these variables.

Table 4

Zero Order Correlations
(See text for symbols)

	<u>DNL</u>	<u>ENTAGE</u>	<u>ENTNL</u>	<u>LOP</u>
DNL	1.000	0.347*	0.510*	0.132**
ENTAGE		1.000	0.258*	-0.090
ENTNL			1.000	0.020

*p .01 **p .05

The most solid relationship in the set of variables is that between entry nutrition level and degree of improvement. Since the biggest nutrition level numbers correspond to the worst status (i.e., 10 is worst, 1 is best) and DNL is derived by subtracting final level from entry level, the interpretation is that the poorest entry status is associated with the most improvement. This relationship holds up even when the other factors are held constant by partial correlation.

Age at entry is also correlated directly with amount of improvement, that is, within the age range of our sample, greater improvement is associated with higher age. Partial correlation analysis demonstrates that some of this effect is spurious and arises from the concomitant association of entry age with entry nutrition level. When this effect is removed, the association weakens (partial = 0.256, p .01), but does not disappear. Hence, we conclude that entry age and improvement are modestly associated.

The most surprising outcome of the correlation analysis is the lack of association between magnitude of outcome and length of participation. One would expect that the longer a child is in

the feeding program the greater would be its improvement--within bounds, of course. Thus it is somewhat surprising that the relationship between improvement and LOP, though statistically significant, is relatively weak. When entry nutrition status and age are controlled for, the relationship strengthens somewhat to significance at $p < 0.01$ (partial = 0.174). One possible interpretation of these results is that, since all the children had been in the program for at least 3 months (in fact, 75 percent had been in 6 months or longer), most of the observed improvement took place early in the period of participation, that is, in the first 6 or 8 months. Because of its obvious programming implications this point should be investigated more thoroughly, but can not be with this data set because it is skewed toward the longer period of participation.

Correlation analysis can give us clues as to associations between variables, but carries no direct implications of causal, predictive relationships. To gain some insight into how these results might be used to predict the outcome of a feeding program we turned to a multiple regression analysis in which change in nutrition level was hypothesized as a function of entry level, age at entry, and length of participation. In a stepwise analysis, the first variable to enter the equation was entry nutrition level, next was entry age, and last, but still statistically significant, was length of participation.

Overall, the equation obtained explains 33.1 percent of the total variance in the system. (In nonstatistical terms, this means that the independent variables tested explain 1/3 of the effect seen, while other factors, unknown and/or unmeasured, account for the other 2/3. While this level of uncertainty may strike the non-social scientist as rather high, experienced analysts would view it as about par for the course in social systems.)

The equation describing the relationship between change in nutrition level and the three independent variables takes the form:

$$DNL = a + b_1(ENTNL) + b_2(ENTAGE) + b_3(LOP)$$

When standardized regression coefficients (the b_p) are developed, they show the relative importance of each variable in predicting the value of DNL. In this sample, the values are: $b_1=0.443$, $b_2=0.246$, $b_3=0.145$. Therefore, ENTNL is about twice as important in predicting the magnitude of change as is ENTAGE, and about three times as important as LOP. In terms of contributing to the accuracy of the prediction of DNL, however, where the entire equation removes 33.1 percent of the inaccuracy, ENTNL by itself removes 26.0 percent, ENTAGE 5.0 percent, and LOP, 2.1 percent. The prediction equation (now using the unstandardized regression coefficients) is:

$$DNL = -3.317 + 0.483(ENTNL) + 0.309(ENTAGE) + 0.554(LOP)$$

Using this equation and associated standard errors, a table can be constructed which both describes the present outcome of the CRS in terms of these four variables and predicts the range of future outcome providing no major changes are made which create substantially different levels of interaction between them. One example of such a change might be a change in the calorie value of the distributed ration. Another might be some change in the health of the children which significantly increases or decreases the bio-utilization of ingested food. Thus, Table 5 should be viewed as today's picture and perhaps tomorrow's, as well, if the CRS program continues along the same path.

Table 5 shows the range of outcome in terms of number of nutrition levels improvement or decline which may be expected with a 68 percent probability for the tabled combinations of entry nutrition level, age (in months) of entry, and length of participation in the program. While there is a 68 percent probability that a child with the stipulated characteristics will fall within the range shown, there is also a 16 percent probability that he/she will do better and 16 percent probability of doing worse.

Table 5

68 Percent Probability Range for DNL as a
Function of ENTNL, ENTAGE, and LOP
(See text for abbreviations)

ENTAGE (mos.)	ENTNL = 5		ENTAGE (mos.)	ENTNL = 8	
	LOP = 9	LOP = 18		LOP = 9	LOP = 18
12	-0.599 - 0.529	-0.284 - 0.844	12	0.670 - 2.158	0.985 - 2.473
24	-0.311 - 0.985	0.004 - 1.660	24	0.958 - 2.614	1.273 - 3.289
36	-0.023 - 1.441	0.292 - 2.116	36	1.246 - 3.070	1.561 - 3.745

- 98 -

REFERENCES

- Jean-Pierre Habicht (1980). "Some Characteristics of Indicators of Nutritional Status for Use in Screening and Surveillance." American Journal of Clinical Nutrition, vol. 33, no. 3, pp. 531-535.
- Reynolds Martorell (1981). "Measuring the Impact of Nutrition Intervention on Physical Growth," Draft Paper delivered at the United Nations University Methodology Workshop, M.I.T.
- Reynolds Martorell et al. (1980). "Improved Nutrition and Its Effects on Anthropometric Indicators of Nutritional Status." Nutrition Reports International, vol. 21, no. 2, pp. 219-220.
- David C. Miller et al. (1977). "Simplified Field Assessment of Nutritional Status in Early Childhood: Practical Suggestions for Developing Countries." Bulletin of the World Health Organization, vol. 55, no. 1.
- Roy I. Miller and David Pyle (1981). Toward a Monitoring and Evaluation System for P.L. 480 Title II Maternal Child Health Programs in India, Community Systems Foundation, Ann Arbor, Michigan.
- David E. Sahn (1980). The Integrated Maternal and Child Nutrition Project in India: Recommendations Based on a Review of Past Experiences, Community Systems Foundation, Ann Arbor, Michigan.
- David E. Sahn and Robert M. Pestronk (1981). A Review of Issues in Nutrition Program Evaluation, A.I.D. Program Evaluation Discussion Paper No. 10, Agency for International Development, Washington, D.C.
- J. W. Thomas and R. M. Hook (1977). Creating Rural Employment, A Manual for Organizing Rural Works Projects, Harvard Institute

for International Development, Cambridge, Massachusetts.

WFP (1976). Interim Evaluation Report on Project Bangladesh 2197/Q,

WFP/CFA: 2/12-A Add. C23, World Food Programme, Rome.

World Bank (1980). World Development Report, 1980, Washington, D.C.

-100-

November 9, 1981

TO: SEE DISTRIBUTION
FROM: Barry Sidman, FVA/PMS
SUBJECT: Principles and Guidelines for the Evaluation of PL 480 Title II Program and Policies

We have asked David Sahn from Community Systems Foundation to review several PL 480 Title II evaluations conducted over the last year. His report, Principles and Guidelines for the Evaluation of PL 480 Title II Programs and Policies, suggests improved methodologies for measuring the impact of the MCH, School Feeding, and Food for Work programs. He will be presenting his conclusions on Thursday, November 19, at 10:00 in the 4th Floor Conference Room, SA-8. We hope you will be able to attend and participate in the discussion. Please feel free to bring other people in your Bureau who are concerned about the evaluation of PL 480 Title II programs.

I look forward to seeing you on November 19, 1981.

DISTRIBUTION:

LA/DP, HJohnson
NE/TECH, BTurner
AFR/DR, JKoehring
ASIA/DP, CJohnson
ASIA,DP, MNorton
ST/N, HSukin
ST/N, SBlumenfeld
PPC/PDPR, JMaguire

ATT. Sahn Report