

PA-ABE-924

ISN 65832

Tony Johnson

WORKING PAPER #2

#2

THOUGHTS ON PERFORMING
ISSUE-ORIENTED STUDIES FOR AID

iate?

Prepared for:

The Agency for International Development
PPC/PDA/CP

By:

Mark Thompson, Ph.D.
John F. Kennedy School of Government
Littauer Center
Cambridge, Mass.

Draft: Not to be cited,
reproduced, or distributed
without the permission of
AID or the author.

TABLE OF CONTENTS

	<u>Page</u>
Introduction and Explanation	1
Operational Summary	3
Supporting Paper A: Terminology and Basic Thoughts for Issue-Oriented Studies	11
Supporting Paper B: Illustrative Derivation of Issues for Study	21
Supporting Paper C: Experience with Issue-Oriented Studies	31
Supporting Paper D: Methodological Alternatives for Issue Study	41

INTRODUCTION AND EXPLANATION

This report addresses the problem posed the Studies Division within the new Office of Program Evaluation and Impact Assessment in AID: how to perform issue-oriented studies. It is based on limited consultation with AID in the late summer of 1977 performed in conjunction with parallel consultation by Professor Robert Klitgaard. Professor Klitgaard has, in a companion piece, described the likely future roles of the Studies Division and its interactions with other units of AID and external bodies. The present report may be read as a sequel to that of Professor Klitgaard. Although cooperation between the two authors was symbiotic and resultant consensus high, the reader will find the two reports distinct in perspective and style, if basically consistent (and, at times, overlapping) in content.

This report is organized about five sections: the first, an operational summary; the latter four, self-contained supporting papers. The four papers present 1) thoughts on issue-oriented studies, ii) illustrative identification of issues for study, iii) experience with issue-oriented studies in other areas of policy, and iv) an assessment of relevant methodologies. Much of the operational summary derives as a selective distillation from the supporting papers. The papers elaborate on many points underlying the recommendations of the summary. Further elaboration is to be found in Professor Klitgaard's paper and its appendices.

To the considerateness and patience of many AID staff members, this report owes much. Particular gratitude is due Donn Block, Allan Hoban, and Herbert Turner.

OPERATIONAL SUMMARY

How to Identify Issues

Recommendation 1: Seek to study issues influencing decisions.

The objective of the Studies Division is to provide useful and needed advice for AID projects based on insights of social science. The long-term viability of the Division requires that its value be rapidly established, which we take to be the near-term goal. The salience, scientific quality, and decision usefulness of initial issue studies are essential to the immediate success of the Division. When the Division is able to produce quickly information sought by the Congress, OMB, or the Administrator of AID and available nowhere else, its stock will soar. This may be achieved by having completed a pertinent study in anticipation of its need or by having rapid study capability. Either case requires an ear to the ground. While Supporting Paper 3 indicates illustratively what researchable issues lurk in annual budget submissions, such methods of issue identification are too ponderous and inexact to be operational. Experience in other areas of government indicates the need for good bureaucratic intelligence gathering (what imminent events will produce what information needs?; on what researchable points will decisions turn?) tempered by judgment (what issues will evaporate?; which are too sensitively bound to value judgments to merit study?). Goals fluctuate over time and vary from observer to observer (pace, GPOI) nor do quantified criteria for their achievement remain constant. In AID, as in domestic social programs, it will be more useful to sense the current

K

priorities of decision makers than to regard permanently codified goal structures (even if arduously built from legislative language). (In the terminology of Supporting Paper A, well-chosen decision-structured issues promise greater return than program-structured issues.)

Recommendation 2: Regard the differential feasibility of issues.

Some issues cannot be researched except at prohibitive cost, others at whatever cost will not lead to clear results, still others will run athwart immobile values. All should be avoided--which will take foresight and judgment.

Recommendation 3: Set pieces may be good (especially in the beginning). Achieving highest decision relevance, as urged in the first recommendation, is not trivial. A reasonable alternative strategy--perhaps simultaneously pursued--is to undertake studies whose outcomes can be confidently anticipated and that will guide practical improvements. As initial foci, such studies have low risk and undeniable gain. Concentration in carrying them out can be on training Division personnel and on setting worthy precedents for scientific quality. The Division head can identify such issues by asking himself "what improvements in project design and implementation can be irrefutably shown by limited application of social science methods?"

Recommendation 4: External publication may enhance prestige.

A further fallback type of study may be one designed to demonstrate scientific competence and relevance if not to influence decisions directly. External publication will achieve salience and will attest to quality. The Division must, however, ensure that such publications

5'

serve the goals of its own viability and do not become overly academic ends in themselves. A limited example of such a study might be investigation of program effects on the rural poor (perhaps a tightly edited book with contributions by many leading minds in the field). Such external indicators of Division success could help to establish it within AID.

How to Study Issues

Recommendation 5: Strive for clarity based on simple and logical organization of facts. This is argued in Supporting Paper D.

Recommendation 6: Avoid expensive and unwieldy methodologies. This is corollary to the fifth recommendation. The additional expense of ideal experimental design, intricate regression techniques, formal planned variation, and large-scale modeling is rarely justified by more accurate information for decisions (although, as indicated in Supporting Paper D, there are exceptions). To these, quasi-experimental design (with due attention to its limitations) is preferable. Many large and sophisticated studies have bombed—a risk that a fledgling office ought not to run. This recommendation does not imply that scientific sophistication should be wholly abandoned. Requisite data should be gathered and digested through appropriate statistical techniques. The goal is to achieve (within resource constraints) maximally convincing documentation of observed effects.

Recommendation 7: Scrounge. This recommendation of World Bank personnel is corollary to the sixth. It will generally be better to make best inferences from whatever evidence and data can be readily obtained than to invoke a mighty methodology demanding vast data-gathering efforts.

Recommendation 8: Methodological development has dividends and dangers. Experience in international development and other policy areas indicates that existing methodology leaves much to be desired. The analytic terrain tentatively assigned the Division is frontier territory. The best researchers within or consulted by the office will inevitably develop their own methodologies as they proceed. These can be useful spinoffs securing external prestige. The dangers in pursuing methodological advances directly are their too academic nature and the high frequency of failure in such efforts.

Recommendation 9: Methodological development seems most promising in i) fashioning more sensitive tools and concepts for social scientific exploration, ii) incorporating subjective opinion or expertise, and iii) modifying analysis to aid communication, reconciliation, and collective decision making. The first category comprises improved questionnaires, survey methods, modes of explanation, and the like. The second includes and should build on the subjective methodologies mentioned in Supporting Paper D. The third is an extension of the second and represents a novel area of unknown potential. Not only does it recognize that better communication can improve analysis but that study processes themselves can foster better communication. One might, for example, engage in simple interactive modeling or gaming based on contemplated development projects and involving spokesmen for different interests.¹

¹Leaders in this field are C.S. Holling of the University of British Columbia and Jose Villegas of Unesco (formerly of Cornell). Computational support may be limited to desk-top or hand-held machines.

7

As the spokesmen see how their concerns are reflected by the models (often by helping to build the models), how these affect other groups, and how other interests impinge on them, they may better reconcile themselves to outcomes. Involvement of social scientists and decision makers from host nations in ongoing studies seems a worthy end in itself.

Recommendation 10: Permit divergent analyses. The best of analysts do not agree on appropriate models, suitable data sources, modes of reasoning, the weights of competing inferences, let alone conclusions. To seek to force immaculate objectivity on analysis is to play Procrustes. Instead, the advocatory nature of analysis should be admitted. One might reasonably present two sides of a policy question and argue for one, but concede that alternate logic or values could reverse the position. This is likely to be the most comfortable and meaningful pose for the office to adopt. On grounds of basic scientific probity, one should make available sufficient evidence that reanalyses based on competing premises may proceed. The greater gain from such policy will accrue not through adversarial battling through to consensus (which indeed may never be achieved) but through the higher standards scientists impose on themselves when anticipating sharp-eyed critiques.

Recommendation 11: Strike a delicate balance between formative and summative evaluations. Ideally, the Division should be seen by missions as a convenient and helpful source of guidance for projects—serving the formative goal of improving them. There will, however, have to be incentives—beyond the engaging natures of Division personnel—to request

its help. There must be sanctions when—from a summative standpoint—project designers and implementers flagrantly flout anthropological or social ecological principles. To advise without overbearing, to effect a mode of mutual cooperation with minimal use of sanctions (both sides have sanctions) will require delicacy. Tact will be prized in recruiting Division personnel.

Recommendation 12: Use trainees. Training programs are excellent sources of cheap labor. The Division should cooperate with them and, from its viewpoint, exploit them. Further advantages are seen in Recommendation 15 below.

Recommendation 13: Cooperate with the DIS without duplicating its functions. Professor Klitzgaard has elaborated this point. The DIS can strengthen contact between the Division and the Missions (constituting an important mode of access to the Division and its work), and will provide research aid. The Division should handle for DIS all non-routine responses involving social science—whether by referring on, brokering, commissioning, or researching. The distinction seems clear enough to minimize turf battles.

Recommendation 14: Be capable of speed. This is based on experience in other policy areas. Decision-structured issues arise without warning and require rapid investigation. Having the capability to produce a respectable study within a month of the issue's emergence will enhance the value of the Division to policy decision makers. Such studies will usefully season longer-term undertakings although an excess of fire-fighting is to be avoided.

How to Promote Utilization

Recommendation 15: Foster demand. The Division may buy time with important books or methodological developments but it will not be secure until its product is needed. Present demand, to be blunt, is poor. Neither Congress, AID/Washington, the Bureaus, nor the Missions are clamoring for issue-oriented studies. If the Congress cared, the others would have to but Congressional incentives lie elsewhere. There is no real constituency for development aid. Specific recommendations below address modes of fostering demand for issue-oriented studies.

Recommendation 16: Instruct. An infrastructure of social scientific sensitivity is vital to the success of the Division. If an enlightened Congressional Committee were suddenly to stress the sociological impact of AID programs, the lack of grounding in social science within the Agency might prevent effective compliance. Sensitization is necessary. Initial documents produced by the Division should (perhaps explicitly) be designed to teach decision makers what social scientific modes of investigation are important and how they might guide policy. This is a further reason why didactic set pieces (see Recommendation 3) are good initial foci. Interaction with training programs—(both the DSP and the IDI) whether in classroom presentations or in field work—will contribute to greater awareness within the Agency for the concerns of issue-oriented studies.²

² A more general area of needed sensitization throughout the federal bureaucracy is in the area of decision making under uncertainty. The policy areas addressed by issue-oriented studies rarely boast definitive resolutions. Competing inferences based on incomplete information are often the only guides. Learning to make best decisions in such situations, giving due weight to the uncertainties involved (but not being daunted by them) is essential. This task of instruction exceeds the capacity of the Studies Division.

Recommendation 17: Present findings clearly and succinctly in the language of intended users. A primary failing of past issue-oriented studies has been their opaqueness. Potential users had to invest too much effort to extract valuable lessons. Congressmen, their staffs, OMB personnel, Assistant Administrators, and Mission members are busy people. They should also be the users of issue-oriented studies. The Division will succeed precisely if these people deem it worth their while to read its studies. In the beginning, it may be well to err on the side of simplicity in presentation (with more technical back-up available to those who wish it). As the Division gradually develops an audience attuned to sophisticated social science, the technical level of presentation may rise.³

Recommendation 18: Retain learning. That AID lacks an institutional memory is often noted. Individuals do not learn the insights of others in distant but similar situations; they forget their own learning or leave with it to take new assignments. The Division should maintain a catalogue of what it has done and has learned. Periodic state-of-the-art reviews should formulate compactly the important lessons of social science disciplines for development projects. These should be addressed to non-technical potential consumers of issue-oriented studies as part of the continuing sensitization of the Agency (as per Recommendation 16).

³We believe the oft-mentioned problem of inadequate dissemination of evaluations to be chimerical. Users do not read evaluative reports because impenetrable style or irrelevant content renders them unhelpful. Pertinent reports pithily written find readers—even without such well-conceived dissemination systems as the DIS. If issue studies distributed by the DIS are not read, their own defects will, most likely, be to blame.

Supporting Paper ATERMINOLOGY AND BASICTHOUGHTS FOR ISSUE-ORIENTED STUDIESFirst Definitions

Issue-oriented studies may be structured by programs or by decisions. Program-structured issues spring naturally from program formulation and pose standard evaluative questions: "Are given programs achieving specified output levels or goal achievement norms?"; "What is the sociological impact of rural electrification"; "Which programs work best in a given country?" The Logical Framework used in current AD evaluations can be extended in a regular way to identify automatically program-structured issues. This extension would adjoin impact dimensions—not goals to be achieved but (possibly diffuse) effects to be monitored—to the GPOI (goal-purpose-output-input) hierarchy.

A decision-structured issue is characterized by its role in decisions and is defined as a consideration or fulcrum on which a decision may turn. Depending on alternative resolutions or judgments of such an issue, different decision outcomes occur.

To identify decision-structured issues, one begins with a notion of the decision faced. Such a decision might, for instance, ponder raising or lowering the support for farmer cooperatives in a given city. Scrutinizing the participants in the decision may reveal that their actions are contingent on the answers to a series of questions:

- 1) Does aid to farmer cooperatives significantly contribute to rural health?

- 2) What population groups benefit most from aid to farmer cooperatives?
- 3) What are the social effects of aid to farmer cooperatives?

Different decision makers may base their actions on any one of these questions or on others. Any question that can elicit different responses affecting program or policy decisions, may be considered a decision-structured issue. Not all such issues will, however, admit to cost-effective means of investigation or evaluation. Those that do are termed evaluable issues.⁴

Further Typology

Issues may be explicitly formulated for decisions—either with a small number of alternative actions in mind or choosing from a large number:

The binary-alternative decision-formulated issue: Should marginal program monies go to transportation or to electrification?⁵

The ternary-alternative decision-formulated issue: Which of three suggested strategies for encouraging land reform ought to be pursued?

The multi-alternative decision-formulated issue: What is the best way of extending credit to small farms (given that one has decided to extend such credit)?; How should one promote an indigenous capacity for agricultural research in developing nations?

Such formulations presume that the respondent can consider all viable alternatives and knows the value structures of involved decision makers

⁴Terminology introduced by J.S. Wholay et al.; "Evaluation: When Is It Really Needed?"; Evaluation; 2:2/1975, pp. 39-93.

⁵Answers here should recognize the importance of context and, accordingly, identify the limits of their own validity.

well enough to determine optimal actions. Rarely is this the case. More often, the evaluator is called on to evaluate one alternative thoroughly. Such evaluations are termed summative⁶ and may result in on-off, scaled, or multi-attribute assessments:

The on-off summative issue (to be answered affirmatively or negatively): Are population programs achieving a predetermined threshold level of acceptance?

The scaled summative issue (to be answered with a single number indicating the overall attractiveness of the alternative⁷): How much value is derived from disaster relief programs?

The multi-attribute summative issue (to be answered by providing information on a range of effects): What are the consequences along many dimensions of interest of programs designed to create and to strengthen intermediate credit institutions?

It is increasingly realized that decision makers do not have modes of valuing alternatives that are readily quantified by an evaluator, or consistent from one decision maker to the next, or even constant over time. As a result, analysts often find it preferable to describe spectra of effects—allowing decision makers to value those effects in their personal, idiosyncratic ways.

Issues formulated as requests for information on certain consequences of a choice or program without valuing those consequences are termed contributory. As with summative issues, distinction may be made among on-off,

⁶ Terminology introduced by Michael Scriven; "The Methodology of Evaluation"; pp. 123-136 in Carol Weiss (ed.); Evaluating Action Programs; Boston: Allyn and Bacon; 1972. The term is, at times, erroneously taken to imply ex post evaluations.

⁷ Benefit-cost and cost-effectiveness ratios are examples.

scaled, and multi-attribute contributory issues. An example of an on-off contributory issue: "Have rural development programs in a country helped to slow migration to the cities?" Contributory issues are termed factual when resolvable by discovering an indisputable fact; and descriptive if resolved by detailed accounts of programs, processes, contexts, policies, or outcomes. While the summative issue comprises all important aspects of a decision alternative, contributory issues concern pieces of information that must be combined with others and appropriately valued to achieve a summative assessment (which in turn must be combined with other summative assessments for decision resolution).

Summative evaluations have been usefully contradistinguished from formative evaluations:⁸ the former tells the overall worth of a program or alternative; the latter tells the best way of implementing a program or alternative. Parallel typology may be extended to issues--differentiated by the type of evaluation they entail. Too much ought not, however, to be made of this distinction. Each type of evaluation depends on the other: summative evaluation presumes that formative evaluation has identified the best version of a program for comparison with other programs; formative evaluation requires summative techniques to determine the values of alternative program strategies. In affective terms, summative evaluations tend to anger program personnel who resent being judged; formative evaluations avoid this emotional impasse by focusing on how to improve programs (without threatening jobs or program funding).

⁸ Scriven, op. cit.

Both summative and formative evaluations may take retrospective or prospective positions. Retrospective summative evaluation attempts to gauge how well a program or strategy has fared in the past. Prospective summative evaluation asks how well a program or alternative may be expected to do in the future. Retrospective formative evaluation examines different implemented alternatives and identifies the most successful. Prospective formative evaluation is more creative. Based on past experiences, on assessments of situation and on innovative reflections, it formulates input combinations and strategies that have greatest promise of achieving future benefit.

All of the issue types above are specific: related to foreseen decisions or to the evaluation of given alternatives. Issues that may influence a broad range of decisions, many not yet specifically anticipated, are termed diffuse. Examples of diffuse issues include: "What are the various demographic, cultural, psychological, economic, hygienic, and medical factors affecting health in rural areas?"; "Under what combinations of conditions does aid to farmer cooperatives prove most beneficial?"

Characteristics of Decision-Structured Issue Study

Standard evaluation techniques and program-structured issue studies are adapted to the recurring needs of management. They are regularly performed to ensure the accountability of input commitments, the diligence of personnel, and the realization of output norms and low-level goals.⁹

⁹In the Logical Framework of AID, but not generally, these are termed "purposes."

The regularity of the structure itself helps to achieve these ends.

The decision-structured issue study (DSIS) is released from strict ties to program structure for the purpose of securing information most useful to decision makers. It is not accountable for coverage of programs, being free to ignore aspects of that structure not of current concern to decision makers, not amenable to feasible study, or entailing exorbitant cost. The decision-structured issue study is instead accountable to obtain information of greatest relevance and value to decision-makers (taking cost into account). From this premise follow many desirable characteristics of the DSIS:

- i) responsiveness to the changing interests of decision makers.
Optimal guidance requires sensitivity to the central concerns of policy debates. These change, often without pattern, and must be responded to. When possible, shifts in priority or perspective should be anticipated. The DSIS is ideally based on a delicate responsiveness to the interests, needs, and moods of decision makers. Such a study should not only base its choice of objects on the concerns of decision makers but should also couch its conclusions in language designed to persuade them.
- ii) addressing the goal and purpose of the goal heirarchy model for evaluation. Issue-oriented studies focus on the higher levels of project objectives—their purposes and goals. When the lower levels are studied, it is primarily to establish their effects on the higher levels. At the higher levels, the very articulation

and quantification of aims has proven most difficult. Goals change over time and vary from person to person. Causation is poorly understood. Instead of berating managers for failing to spell out criteria capturing purposes and goals, one should recognize the fundamental intractability of that task. Studies oriented to issues should plumb the relationships among outputs, purposes, goals, and priorities, and help to explain them to decision makers.

- iii) complementing project evaluations. Current project evaluations best relate inputs to immediate outputs. They most often cannot—due to the limited nature of projects—quantify movement toward higher-level objectives. This role should be played by issue-oriented studies. As an example, a project evaluation might identify direct gains to specific farmers achieved by a rural assistance project. Issue-oriented studies would next be invoked to determine the extent to which such immediate benefits contributed toward such higher-level goals as alleviation of rural poverty or fostering agricultural security. Note that a single issue-oriented study (best based on many sites to understand the importance of contextual differences) could be used to interpret the significance of any number of projects attaining the direct gains.
- iv) focusing on relevance to decisions. The informational worth of evaluation derives from its ability to aid decisions. Most useful direction of evaluation studies requires attention not

only to the current interests of decision makers but also to the nature of impending decisions. Issues likely to sway many decisions should, ceteris paribus, have priority for investigation. In like vein, concerns of external validity (generalizability) should guide the choice of study object.

- v) Regarding feasibility. Decision-structured issue studies enjoy flexibility in task formulation. This flexibility should heed not only decision relevance but also the differential feasibility of evaluations. When relative infeasibility is indicated by either high costs or low expected precision, such studies should be forgone in favor of more tractable alternatives. Evaluability assessments¹⁰ may be invoked to determine the likelihood of evaluation success. While tight experimental design can often be maintained for measurement of direct goals, issue-oriented studies often embrace such large social scope as to preclude classical control. In such cases, quasi-experimental design as spelled out by Campbell and Stanley¹¹ may represent the most promising approach.

Rationale for Issue Study

The justification for studying either program-structured or decision-structured issues is the promise of improving decision making. Decision

¹⁰ See Whorley, op. cit.

¹¹ D.F. Campbell and J.C. Stanley; Experimental and Quasi-Experimental Design for Research; Chicago: Rand McNally; 1966.

theory¹² has defined the prior information value (PIV) of a study as the likely gain realized through improving decisions. Loosely, this is, for any study, the product of the likelihood that a decision will be altered by the study and of the average gain achieved when the decision is affected. Study cost is taken into account in the net prior information value which is the PIV net of costs.

This decision-theoretic formulation enables useful generalizations on study initiation: in focusing only on the informational worth of studies, the net PIV is the bottom line. A study with negative net PIV ought usually not to be commissioned. In such cases, costs exceed expected gains. Exceptions to this rule occur when studies promise significant non-informational benefits (such as stimulating personnel). Among alternative studies with positive net PIV, the best choice is that with highest net PIV (assuming that non-informational concerns do not weigh heavily).

That one optimally maximizes net PIV—probability of changing decisions times average value of change less cost—aids in study selection. Studies should, to maximize the chance of affecting decisions, address decision makers in terms that persuade them. Decision-making groups with quantitative orientation should be presented numbers and statistical verifications. Those legally minded should receive syllogistic reasoning from premises. Conversely, studies based wholly on arguments unlikely to be understood or heeded by decision makers have little decision value.

¹²The general best reference is Howard Raiffa; Decision Analysis; Reading, Massachusetts: Addison-Wesley; 1968. Specific application to decisions is given in M. Thompson; Evaluation for Decision in Social Programmes; Westmead, England: D.C. Heath; 1975.

Yet, as the criterion of net PIV reminds us, influencing decisions is not the lone component of study value. Only insofar as that influence improves decisions are benefits achieved. In addition to being persuasive, studies of higher value must be accurate and right. The cost term in net PIV indicates that such study benefits must themselves be traded off against expenses borne.

For specific issues, net PIV is—because target decisions are identified—more readily estimated. For diffuse issues, possible affected decisions are more hazily perceived, making net PIV dependent on more general notions of issue relevance. In the latter case, generalizability—the confidence one has in extrapolating study insights to different contexts and decisions—is a prime concern.

For either specific or diffuse issues, a common set of questions should be posed in commissioning a study:

- 1) should important decisions be affected by investigation or resolution of an issue?
- 2) can a study of the issue be instituted for reasonable cost?
- 3) would such a study contribute substantially to understanding the issue?
- 4) would decision makers pay attention to such a study and allow their actions to be influenced by it?¹³
- 5) if decisions can be partially determined by such a study, can one be confident that this influence will be beneficial?

With affirmative answers to these questions, an issue study should be initiated. In choosing among alternative studies, the principal concern should usually be to maximize the net PIV.

¹³An affirmative answer implies that due attention is paid to the factor of timing.

Supporting Paper 3ILLUSTRATIVE DERIVATIONOF ISSUES FOR STUDYMethods

To show how study topics might be selected and to get a sense of current informational needs, we consider issues identified in AID staff reviews for the FY 1979 budget. Because budget review follows the format of existing program divisions, many issues so isolated are program-structured. Since our criteria for recommending further investigation will be decision relevance, all issues so designated will be decision-structured. These criteria are spelled out in the five questions given on page 20. This review is undertaken without the advantage of extensive inside operational knowledge (which issues are hot, which dead, which would be blocked, which would fan flames). Such savvy should be applied along with the general criteria here illustrated to single out specific issues for actual study.

The documentary base for this review consists of a series of memoranda prepared for budgetary review in August, 1977, by the Bureaus of PPC and SER. The memoranda bore the paraphs of the two Bureau heads and identified issues abstracted for PPC/SER discussion from the budget submissions of the various Bureaus and Offices within AID. One document touching on a broad range of questions is that for the African Bureau. We have summarized many of the issues it raises and scrutinize their suitability as objects for formal study.¹⁴

¹⁴Wording within quotation marks comes directly from the African Budget Review memorandum of 17 August, 1977. Issues not quoted directly have been paraphrased for conciseness.

Issues Identified in the African Budget Review

1. Can the large increase proposed (\$412 million for FY 1979, versus \$258 million for FY 1978) be effectively implemented and managed? (on-off, prospective, contributory)

Budgetary and program decisions should benefit from an answer to this question. The most appropriate type of study is moot. Managerial expertise with understanding of the particular context is required. It seems likely that contextual familiarity is so important that general managerial acumen will not be especially useful. The persons involved in this budgetary decision can probably answer the questions better than any analysts they might engage. A review of past experiences in rapid program expansion—perhaps critiqued by specialists in multi-project management—might usefully counsel these decision makers. In sum, however, it appears that a study of this issue—especially a large study—would not sufficiently add to the understanding of the decision makers to justify its costs.

2. "How many contractors will be required in the field in 1979 to implement the proposed program?" (scaled, prospective, contributory)

Such information should aid staffing decisions. When combined with information on the pool of potential contractors, this issue will help to determine budgetary choices. Here again, a basic managerial sense of requirements and options seems required. As this will most likely repose with AFR Bureau managers of the present or recent past, a formal study appears contraindicated. Preferred will be simple direct commentary communicated by such persons. It is possible, however, that there is more here than meets the eye. An important, latent issue might be the fall-off

in program efficiency when too few contractors or those of too low quality are employed. Whether in fact a study could be profitably addressed to this related issue should be judged by those most familiar with these decision problems.

3. "Does the modest annual rate of increase of 10% after 1979 reflect a belief by AFR that the 1979 program will have pretty nearly saturated Africa?" (on-off, factual, contributory)

As couched, this issue is a question of fact to be resolved by a telephone conversation rather than a study. As with 2 above, more fundamental issues lurk beneath the surface: what does it mean to "have pretty nearly saturated" a continent?; how does the average value realized per marginal project vary as the number of projects implemented in a region grows? The latter question could be the focus for a full-fledged study.

4. Given program size, dispersion, and operational constraints, "should we not concentrate on those countries in which the most promising conditions exist for a successful program and in which programs of reasonable size can be implemented?" (binary-alternative, decision-formulated)

This decision could give rise to an elaborate issue study estimating the probable impact of such suggested concentration in comparison to the more diffused budget proposal. The study would identify the various impacts realized by more and less consolidated programs and would estimate the differential values of those impacts. Managerial experts and personnel familiar with the program would be most adept at gauging direct impacts. Economists, anthropologists, political scientists, demographers, and social psychologists might be among the relevant specialists for assessing

indirect impacts and for valuing the different effects. Reliable advice tells us, however, that such a study would be unwise¹⁵ since decision makers on program concentration would not be swayed by it. This underlines the importance of sensitively grasping decision-making attitudes if one is to avoid commissioning studies that cannot affect policy. We note also that the identification of issues is inevitably subjective. The author of the AFR memo was intrigued with the idea of managerial concentration and found it a concern of importance for five separate missions. Another observer might have abstracted and stressed a different series of issues.

5. "Are steps being taken in project design and implementation to assure that AID activity is not aggravating the human rights problem (six nations are identified) or handing support to those responsible for the situation?" (on-off, factual, retrospective, and prospective)

Mention of design and implementation indicate that this is not only a direct budgetary decision to allocate or to withhold monies on the basis of their human rights effects. Budgetary review may also be used to give incentives for appropriate decision making in project planning and management. The factual issue formulated requires no more than one reassuring communication. Researchable sibling issues may, however, be posed: what effects do AID programs have on human rights?; what managerial policy would best serve U.S. concerns in this regard? Research on the latter question seems impractical. Extensive and expensive communication with highest-level policy makers would, on the one hand, be

¹⁵ A "non-starter."

entailed. On the other, on-site research seems prohibitively sensitive. The former issue might be addressed with a retrospective study examining past patterns, incidence, and effects of AID support in nations with human rights problems. Useful data resources would be centrally-held files and querying of AID personnel. The disciplinary perspectives of political science and sociology would be valuable in such a study.

6. What is the internationally coordinated development strategy for the Sahel? (descriptive, factual, prospective)

This summarizes a long series of questions raised with regard to the international strategy and U.S. participation in it. Some questions relate to vital concerns (e.g., absorptive capacity for funds, infrastructure support) that may be foci for worthwhile studies. It is not clear, however, if and how such information would be used by AID decision makers. If policy is effectively set by international consortium and not influenced by AID studies, they should not be performed.

7. "What is our approach to food production projects in countries whose price and marketing policies act as a disincentive to food production by the small farmer and herder and which pursue policies which promote cash crop production at the expense (sic) of food production?" (descriptive, factual, prospective)

The posing of this factual question indicates the potential value of an issue study investigating what our approach "ought to be" to these projects. Agricultural science, farm marketing, economics, and anthropology are just a few of the disciplines to be brought to bear. Such a study should address the background conditions influencing the differential effectiveness of projects.

8. "Both (Rwanda and Burundi) have serious population problems and neither have population policies or programs."

This declarative description of a problem may be reformulated as an issue for study: what should U.S. policy be toward nations not moving to solve their population problems? This is a subject many times broached by AID studies¹⁶ and those of other donor organizations. Factors that might justify mounting yet another study include: context specificity—appropriate policies depend on unique national characteristics that may necessitate many separate studies; changing contexts—these characteristics vary, as do the priorities of U.S. policy; and disciplinary coverage—anthropological understanding, for one example, may have been inadequately represented in certain past studies.

9. Given the economic deterioration and lack of effective political leadership in Ghana and Sierra Leone, should we not consider curtailing or terminating these programs? (trinary-alternative, decision-formulated)

This question can only be answered by those with an embracing grasp of policy imperatives and options. As such, it comprises an inappropriate focus for an issue study. Useful issues contributory to this question include: what projects in such contexts are most effective? What is the worth of such projects vis-à-vis projects in better circumstances?

10. To what extent will AID development programs in Kenya and Chad be handicapped by most government emphasis on military expenditures? (scaled, prospective, contributory)

This question may be illuminated by the same types of contributory issues given in 9 above.

¹⁶For example, in the 1970 Spring Review.

11. "Is it reasonable to assume that we could exert such control over the sector loan local currency proceeds that they would indeed go for support of our activities even though the (Government of Zaire) continues to attach highest priority to assistance to urban population?" (on-off, prospective, contributory)

As couched, this is a single-nation issue perhaps best handled (possibly by commissioning a study) by the local Mission. Should this problem extend to other nations and admit to inquiry generalizable across national differences, a more broadly based study of those and related issues (e.g., what forms of loan control exert most leverage with least pain?) might be initiated by the Bureau or by AID/W.

12. "One of the major objectives of the (Sahel Development Program) is an increased food production and income for the population living in the Sahelian zones. On the other hand, the cost-benefit ratio of our assistance projects in food production is likely to be much greater in the southern, higher-rainfall area. How does the FY 1979 program deal with this issue?" (descriptive, factual, prospective)

This is another issue set as a point of factual investigation but suggesting the more fundamental question of determining appropriate strategy. Summative issue studies of aid in the North and South (e.g., of Chad, for which this issue was raised) would assist this determination. The appropriate organizational unit to commission such a study would depend on the extent and generalizability of the problem. The phraseology used suggests a basic misunderstanding of cost-benefit analysis which, properly applied, does not exclude valid distributional or humanitarian

concerns. Introducing cost-effectiveness analysis might avert this misconception.

13. How should one, in Mali and Mauritania, design programs to aid the rural poor rather than the better-to-do families? (multi-alternative, decision-formulated)

This wording presupposes that this type of aid is the more desirable. If moot, this would be an excellent object for an issue study. The actual formulation might be investigated through a review of past projects and through the use of social science to understand the economic, cultural, and demographic mechanisms at work.

Overview and Conclusions

We have summarized and discussed above the issues raised in but one segment of FY 1979 budget review. Other review documents picked out different issues: strategies for accommodating concern for the poor, the general problem of fitting Mission-generated projects with Congressional guidelines on program types, inquiry into the likely effects of marginal changes in funding or personnel, questions on the best way of coordinating AID programs with those of other donor organizations. These differences bespeak both the programmatic perspectives of the various Bureaus and Offices and the divergent backgrounds of document authors. Notwithstanding these variations, the issues raised in the AFR document, give a representative flavor of those found in the other submissions.

In reflecting on the whole set of issues raised, we are struck by conflicting impressions. We note first the superficiality of many questions:

what is the Mission's justification for a proposed increase?; how many contractors are needed?; what is our policy in an area? Those questions demand curt factual answers and yet, as argued at many points above, more fundamental issues (how should competing budget priorities be traded off?; how should optimal policy be formulated?) hide within them. It nevertheless seems difficult to fault the budget reviews for this superficiality. The time horizon of budget review does not permit considered contemplation of deep issues. But this does not imply that these issues should not be addressed anywhere in the Agency.

If, however, such issues are not raised at the budget reviews, they will not be reflected in policy. As in most organizations, budget determination in AID is the process node to which all others refer, project initiation, design, and implementation can be instructed to incorporate certain principles but will consistently adhere only to those enforced by budget allocations.

But how can the delicate and painstaking process of policy formulation be meshed with the frenzied pace of budget preparation and review? A number of issues abstracted are couched in terms of "what is our policy...?" while we have suggested that, posed as "what should our policy be...?", they would usefully guide issue studies. Yet neither of these questions should have to be asked at the budget review. Instead, issue studies and policy reviews carried on over the course of the year should ensure that budgetary decision makers have a continuing sense of these priorities. That sense should be why they are budgetary decision makers. The finer insights of

issue studies (what types of mechanisms for rural credit work best in given regions) should be supplied by staff work.

Yet none of this will work if incentives are not reinforced by highest budgetary actions. The message transmitted by Congressional appropriators is that certain amounts of monies should be spent according to set guidelines that fluctuate annually. It is more important to the Congress that the appropriated amounts of monies be spent (on projects carefully worded to fit the guidelines) than that proof of their efficiency in achieving policy goals be given. This attitude inevitably affects all levels of decision making: OMB, central administration of AID, the Bureaus, and the Missions. Quality is traded for quantity. Until this sense of Congressional imperatives changes, it will be hard to alter significantly the project-approval process or even to blame the many levels of decision making for being responsive to existing incentive structures.

It nevertheless behooves all managers in the organizational structure to promote project quality by all possible means. A good theoretical way of accomplishing this is to follow precisely the strategy of AID: to require separate steps in project planning (PID's, PP's) enabling quality reviews. Unfortunately, this can be thwarted by lower-level management playing on the overriding priority of funding the requisite volume of projects. By delaying submission until the budget schedule permits but cursory review and revision, projects of marginal quality achieve approval.

Supporting Paper C

EXPERIENCE WITH ISSUE-ORIENTED STUDIES

General Experience

Many areas of policy have experimented--some hesitantly and against entrenched instincts, others boldly--with issue-oriented studies. The supporting paper below summarizes--necessarily in a selective, impressionistic way--portions of that experience. As will be seen, approaches, methods, and results have varied broadly. If one were, however, to abstract a general historical pattern, it is the following:

Stage 1: little formal analysis;

Stage 2: with the emphasis on planning, programming, and budgeting systems and program evaluation in the 1960's, attempts were made to quantify goal achievement and thereby to measure numerically program success; and

Stage 3: with the frustration of Stage 2, more limited, often issue-oriented, studies were undertaken, researching selected aspects of program impact. These often reflect awareness that the cumulative impact of diffuse program effects must be appropriately measured and weighted for consideration along with primary program goals. We are now in Stage 3.

We shall below examine specific variations on this theme and identify salient lessons (as, with our imperfect foresight, they now appear).

Bio-Medical Research

The ambitious goal entertained in 1970 for the prospective evaluation of bio-medical research was to estimate in economic and social terms the anticipated impact of proposed projects.¹⁷ Such assessment was intended not to replace scientific peer review but to augment it vitally. Largely because the ultimate impacts of the research were contingent on too many uncertain factors, these hopes were disappointed. As a result, the NIH evaluation plans currently call for more limited studies of relevant aspects of bio-medical research impact. Some such tasks are purely economic (e.g., developing a price deflator for bio-medical research and development); others seek the multifaceted, interdisciplinary perspective of an issue-oriented study (e.g., "The Impact of Federal Policies on Bioscience Research: Scientists' Careers, Productivity, Development of Disciplines"). The parallels with AID evaluation processes are evident (the prospective estimation of diffuse, multi-attribute impact), yet important structural differences are present. Perhaps the most significant difference is the intense competition for the scarce monies of bio-medical research. Experience with studies oriented to multiple issues is presently too recent and incomplete to allow firm judgments on their policy impact.

Education

Educational evaluation was long thought to have the advantage over many social policy areas of a consensus measure: cognitive growth. The

¹⁷ See, for example, R.S. Cutler and V.A. Martino; Project SOAP: Bio-medical Research Relevance Criteria; Washington, D.C.: National Bureau of Standards; NBS Report 10-423; 1971.

33

uproar that followed the Westinghouse Learning Corporation evaluation of the Head Start Program focused attention on several other aspects of importance. Among these were affective growth, parental involvement, physical health, and social integration. More broadly pitched evaluations—composed of many issue-oriented components—were accordingly prescribed for the Follow Through Program¹⁸ and for the Elementary and Secondary Education Act, Title I.¹⁹ Due to the inability to maintain experimental conditions, to the absence of accepted measures for the non-cognitive goals of education, and to affective school-level resistance to reporting requirements, these evaluations were also found disappointing. To date, outstanding prototypes for issue-oriented studies in education have not been achieved.

Recent high-level officials in educational evaluation²⁰ feel that the issues on which decisions depend change rapidly and thus require alert identification and rapid investigation. A premium is placed on speed. Decisions have been significantly influenced by simple transfer impact studies (e.g., who gets how much more money under current and proposed formulae) and minimally affected by the numerous sophisticated regression studies (which often attempt to capture the interrelationships of economic and social effects).

¹⁸ A critique is given by R. Elmore; "Lessons From Follow Through"; Policy Analysis; 1:3 (1975); pp. 459-485.

¹⁹ See M. McLaughlin; Evaluation and Reform; Cambridge, Massachusetts: Ballinger; 1975.

²⁰ This paragraph is based on a number of interviews.

Energy

Whereas energy development decisions in the past have been addressed by benefit-cost models,²¹ the point has now been driven home that no single scale—such as a benefit-cost ratio—can adequately reflect the many relevant considerations in energy policy. As a result, separate substudies are now addressed to such issues as economic effects, environmental impact, safety, nuclear security, and the overall flexibility of policy.²² It is instructive that the relative importance of these issues has over time fluctuated greatly in the minds of decision makers (President Carter's recent decision to cancel the Clinch River Project being largely based on the plutonium issue).

Health

Somewhat later than other policy fields, health analysis turned to benefit-cost models. Rapid perception that full agreement on the monetary value of a human life will never be reached has led to current emphasis on cost-effectiveness analysis. In this framework, enhancements and extensions of life are expressed in terms of dollars per quality-adjusted life year (QALY). The single largest experiment in the health area—the National Health Insurance Study—focuses on the costing of alternative health care provisions and on resultant health status—with such issues as appropriate system structure being neglected. In technology assessments for health care, disparate issues are addressed with increasing

²¹Reviewed and criticized by Thomas Cochran; The Liquid Metal Fast Breeder Reactor; Washington, D.C.: Resources for the Future; 1974.

²²For example, S.M. Keeny, Jr.; Nuclear Power Issues and Choices; Cambridge, Massachusetts: Ballinger; 1977.

32

frequency. A case in point is the prospective analysis for policy on amniocentesis which speaks to such varied concerns as benefit estimation from the reduction of mental retardation, risk assessment, ethical and legal aspects, and effects on the societal gene pool.²³ It seems tacitly agreed that here, as in such other policy questions as that of malpractice, the appropriate conceptual approach is through multiple issues.

Interior

The Bureau of Reclamation and the Corps of Engineers did much to develop benefit-cost analysis as a tool for deciding on and justifying their projects. In recent years these analyses have come under severe fire. Some criticisms—e.g., that the average cost overrun is 72 percent²⁴—do not pertain to the conceptual framework but to its application. Others have been met by extending the scope of the analysis to embrace secondary and tertiary effects—for instance, income loss to farmers who are hurt by competition from others farming reclaimed land.²⁵ As these refinements are made, the heavy influence of politics on analysis is evident in the extensive use of grandfather clauses and in the byzantine computation of the time discount rate. Still other considerations—environmental effects and income redistribution—are not conveniently handled by benefit-cost analysis. A recent official in the Department of the Interior admits the defects of benefit-cost analysis but asserts that

²³ See Aubrey Milunsky; The Prevention of Genetic Disease and Mental Retardation; Philadelphia: W.B. Saunders; 1975.

²⁴ R. Haveman; The Economic Performance of Public Investments; Baltimore: Johns Hopkins Press; 1972.

²⁵ See S.H. Hanke and R.A. Walker; "Benefit-Cost Analysis Reconsidered: An Evaluation of the Mid-State Project"; Water Resources Research; 10:5(1974); pp. 898-908.

it has been useful in weeding out the worst of projects. It appears that reliance on benefit-cost analysis in the Department of the Interior will be increasingly augmented by studies of separate issues.

International Development

AID. The details of past issue-oriented studies are covered in the appendices to Professor Klitzgaard's report. We restrict ourselves here to a brief summative impression parallel to those of other policy areas, focusing in particular on the spring reviews. The typical texture of the issue-oriented report in AID is narrative: how many acres of land in a given country were in one year sown with what. Occasional forays into sophisticated analysis—e.g., simulation models, linear programming, multivariate regression—were undertaken.²⁶ In general, the material seems to suffer not from lack of sophistication (the impressive techniques occasionally applied did not greatly further insight) but from indigestibility. There was too much in dead weight; no effective integration and compaction of the vast information mass was achieved; lessons for future guidance were fragmented needles in haystacks—first to be found, then pieced together. Attesting to this state of affairs are findings on the utilization of economic research²⁷ which revealed that few enough (31%) of potential study users were aware of them, that fewer (5%) read them carefully, and that the studies exerted minimal influence on policy (despite general concession that the professional quality was high). Laudable steps have only recently been taken to achieve practical policy guidance

²⁶All these are found in the 1973 Spring Review on Small Farmer Credit.

²⁷"Utilization of Economic Research"; AID Evaluation Paper #4; October, 1971.

by means of succinct, readable summaries. It is to be hoped that future issue studies will extrapolate this effort.

World Bank. Ex post reviews in the World Bank differ from those of AID in their preeminent concern with rate-of-return measurements. While recent years have increased emphasis on alleviating poverty and on meshing growth with equity, the primary analyst has remained the economist. Sociological, demographic, and other dimensions of social science come to bear through the eyes and pens of economists. When tough and broad summative assessments have been undertaken (as, for instance, in reviewing the entire Colombian program) project-level resentment has been high. The Bank relies heavily on extramural consultants contacted as individuals. It has found minimal assistance for its studies in the academic literature. While many aspects of the Bank's ex post evaluation program are appealing, it is difficult—even for the evaluators—to identify its impact on decisions.

Justice

Multiple issues have been considered in many recent studies within the law enforcement field—notably in the areas of bail bond posting²⁸ and drug control.²⁹ Arrest records and other simple indices of goal attainment have been found suspect (most strikingly in methadone maintenance experiments where lowered arrest rates have reflected improved ability to avoid arrest rather than less crime) and are augmented with sociological

²⁸ The bail bond experiment of the Vera Institute.

²⁹ See M.H. Moore: Buy and Bust: The Effective Regulation of an Illicit Market in Heroin: Lexington, Massachusetts: D.C. Heath, 1977.

and economic gauges of program effect. For purposes of bail bond posting, the concept of the strength of community ties has been found an excellent predictor of bail-jumping behavior. In the best-known of law enforcement experiments—the Kansas City Preventive Patrol Study—no fewer than 173 separate indices (each reported and tested for statistical significance) gauged sociological, criminological, and economic effects.³⁰ This experiment unfortunately was poorly controlled (with little difference in police sightings across the normal, reactive, and proactive beats) and thus forfeited the potential utility of the extensive battery of measures.

Urban Policy

One of the first attempts to obtain multiple unfettered impressions of program effects was in the Community Action Program where a number of contractors were given leave to evaluate whatever aspects of program performance they felt most relevant. Numerous issues received attention in a disjointed way. The contractors never were given a sense of what decision makers considered most important and produced, in consequence, diffuse and disappointing reports.

Multiple issues in housing policy were addressed in the 1973 report of the Department of Housing and Urban Development, Housing in the Seventies. Based on an elaborate econometric model, the study estimated equity effects, three separate measures of impact, and five different measures of efficiency. While presenting a sophisticated conceptual format, the study used many

³⁰ George L. Kelling et al.; The Kansas City Preventive Patrol Experiment: A Summary Report; Washington, D.C.: Policy Foundation; 1974.

rudimentary tools. The proxy measure for societal impact—better health, greater family stability, better school performance—was the price fluctuation in neighborhoods into which subsidized housing was introduced. Democratic factions in the Congress took the report as an attack on the policy (it was used to justify a moratorium in the subsidy programs) and counterattacked with A Critique of "Housing in the Seventies." The ensuing political crossfire obscured the many points of merit in each study in their approaches to the issues.

Welfare

The most notable recent study pertaining to welfare policy is the New Jersey Income Maintenance Experiment.³¹ It originated with the narrow focus of measuring any changes in work behavior under a program of negative income tax but soon was caught up in the multi-issue debate surrounding the proposed Family Assistance Program. Congress pressed for intermediate results, then criticized their intermediacy. In the initial testimony of study leaders before the Congress, an objective, statistical presentation was not understood and they were asked to speak as advocates to be comprehensible. As soon as they took advocacy positions (in favor of extensive welfare support), adversaries were able effectively to eject study findings from the policy debate on the grounds of transparent bias. Since these findings were primarily technical—and hence less subject to partisan manipulation—one may suspect that political foes would yet more easily press home the charge of prejudice had the study focused on more qualitative social issues.

³¹ D. Kershaw and J. Fair; The New Jersey Income Maintenance Experiment—Volume I: Operations Surveys and Administration; New York City: Academic Press, 1974.

Lessons

Supposed lessons have in the past been so often erroneously drawn in the policy areas above that further inference must proceed with caution. As; however, we have argued the necessity of abstracting in a comprehensible way the gist of experience, the attempt must be made. Succinctly, the main points seem:

1. That single-scale measures (rates-of-return, benefit-cost ratios), for all their advantages of simplicity, are being found increasingly inadequate and hence promote the approach through multiple issue studies.
2. That measures of sociological impact are difficult to formulate and to apply, and have frequently been found wanting (often as philosophical disagreement takes the form of methodological challenge).
3. That the mode of presentation may vitally affect the influence of the issue-oriented study.
4. That issue studies are most efficiently accessed by contacting knowledgeable persons. Supplementation of this search mode may be achieved through standard bibliographic methods--including computerized key-word searches.
5. That the importance of issues in policy deliberations varies over time and from person to person thus necessitating firm guidance of issue-oriented studies to ensure their focus on aspects of highest relevance.
6. That political repercussions are not to be escaped: political considerations may limit the usefulness of studying certain issues; analysts in sensitive areas should prepare themselves for scathing attacks on all aspects of their work; the more relevant and potentially more valuable studies will occasion the greatest political heat.

7. That the most sophisticated of modern analytic techniques have but rarely proven more useful to decision making than reasoning based on relevant facts organized by basic commonsense.

42

Supporting Paper D

METHODOLOGICAL ALTERNATIVES FOR ISSUE STUDY

Ideal Experimental Design

Sparked by the prestigious, sweeping review of Gilbert, Light, and Mostaller,³² many evaluation experts urge adherence to principles of classical experimental design. Such design requires randomization of subjects and controls (best stratified) and tight supervision to ensure that specified treatments are followed and do not spill over to controls. Ideally, treatment effects are quantitatively measured but qualitative measurements and statistical analyses are also possible. Precision increases with size. This outlook is reflected in the implementation of such massive and extensive social experiments as those measuring behavior under national health insurance and under income maintenance.

Ideal experimental design excludes all confounding effects that spring from noncomparability of treatments and controls. When such effects are ruled out, overall measurement accuracy is enhanced. An important concomitant result is that spurious findings cannot be consciously or inadvertently engendered by such confounding. In particular, selective reporting or publications are less likely to lead to false conclusions. Gilbert, Light, and Mostaller argue, however, that the other side of the coin is the more important: that statistically significant effects discovered by randomized field trials are much more likely to persuade critical decision makers. This derives from the fact that all valid

³²"Assessing Social Innovations: An Empirical Base for Policy"; pp. 39-193 in C.A. Bennett and A.A. Lumsdaine (eds.); Evaluation and Experiment; New York City: Academic Press; 1975.

43

effects discovered in non-randomized trials can be attributed, by those reluctant to believe, to differences between treateds and controls.

The drawbacks to tight experimental designs are their residual imprecision and their cost. While selection biases are precluded, invalid findings may be induced by defects of external validity (e.g., the Hawthorne Effect) or through statistical bad luck (effects valid at the statistical significance level of .05 occur once in twenty times by pure chance). To reap its statistical advantages, randomization must be performed over units large enough to avert contamination. To evaluate city-wide policies, cities may—at prohibitive cost—have to be randomized.

Reliance on randomized, controlled trials for evaluation within AID is further handicapped by concerns of generalizability. Even if one through rigorous expensive experiment establishes that a program achieves satisfactory results in one set of contexts, there is no assurance that the results will obtain in extrapolated circumstances. Nevertheless, one can imagine instances when ideal experimental designs would be useful for AID programs:

- i. it is desired to know which of two distinct strategies for aiding the rural poor puts a higher percentage of monies spent directly into their hands; and
- ii. a disease control program is epidemiologically effective but may carry risk for inoculees and, if so, should be terminated.

The advisability of experimental design is contingent on circumstances:

- i. an impending sharp decision (choice of program, to inoculate or not). While theoretically one also wants precise information

- for less edged decisions (determining appropriate budget levels), the payoff to sounder knowledge (which is the justification for the additional expense of tight experiments) is more striking when decision alternatives are distinct;
- ii. an accepted measure of experimental effects (percentage of money going to the poor, statistical risk). Without such a measure (when goals are money or fuzzy), the gains of greater experimental precision are lost amid other uncertainties;
 - iii. no ethical costs. Randomization may be unethical if subjects are assigned to treatments strongly suspected to be inferior. This consideration may preclude or curtail testing an inoculation program (if the presumption of risk is strong); and
 - iv. generalizability. Information value is directly related to generalizability—which, in the two examples given may be at issue. Extensive monies should not go to experiments whose findings are thought to be valid in but limited settings.

Quasi-experiments

For a number of reasons, tight randomized control cannot be exercised over many of AID's experimental programs. These reasons include expense, ethical considerations, imperfect anticipation of informational needs, lack of staff, and inability to maintain control in the field.

Campbell³⁴ has suggested that the appropriate strategy is then to

³³ See M. Weinstein; "Allocation of Subjects in Medical Experiments"; New England Journal of Medicine 291: (1974); pp. 1273-1285.

³⁴ D.T. Campbell; "Reforms as Experiments"; pp. 137-223 in Carol Weiss (ed.); op. cit.; or Campbell and Stanley; op. cit.

abstract maximally useful information from less well controlled situations. This may be done by intelligent, commonsensical inquiry into data provided--whether on the differences observed in program subjects versus controls or on the changes in subjects over time. Such inquiry may also be based on multivariate regression methodology. In either case, the term "quasi-experimental design" is applied.

The defects of quasi-experimental design in contrast to ideal experimental design are the much-bruited threats of internal validity.³⁵ These boil down to selection problems (even with sophisticated covariance adjustments, one can never be sure of appropriately comparing treated to controls when selection has been non-random) and problems of concomitant effects (one can never be confident in time series designs that all effects simultaneous to program initiation have been captured by the analysis). Quasi-experimental designs have at times been pushed beyond the reasonable limits suggested by these effects. These abuses have recently led Campbell himself to wonder aloud publicly whether his exposition of quasi-experimental designs had not done more harm than good.

While recognizing the shortcomings of quasi-experiments, we see no appealing substitute for them in the uncontrollable settings of most AID programs. We find, moreover, that most of the effectively persuasive AID documents are based on quasi-experimental presentations. Indeed, we feel that the greatest potential improvement for AID evaluations derives not from the wholesale imposition of ideal experimental design, nor from the introduction of flashy new methods, but from strengthening

³⁵ Ibid.

the interpretation and management of quasi-experiments. We shall return to this point.

Search for Explanations

Complementary to the confirmatory and measurement approaches of statistics are the exploratory and the descriptive. Setting out without firm preconceptions of underlying models, the exploratory statistician searches aggressively for explanations of phenomena. He seeks combinations of variables that are closely related and may hence be collapsed (reducing problems of multi-collinearity). He looks for interaction effects—special fortuitous combinations of variables with significant impact on program outcomes. He may engage in feedforward: determining the targets of data collection by what he learns in the initial stages of investigation. Cluster analysis, factor analysis, path analysis, and principal components are primary methodological tools used in exploratory and descriptive data analysis.

A drawback to exploratory statistics is that, if one looks long enough, one always finds something. This may, however, be a random statistical artifact rather than a valid effect. Similarly, if one invents enough different ways for examining the same phenomenon, one perspective will eventually lend credence to prior beliefs and biases. Significance testing for exploratory analysis is in its infancy but two related approaches seem to hold most promise: withholding a portion of available data from model development, then using it for confirmatory testing; evaluating models on the basis of their success and prediction.

The multivariate methods of descriptive and exploratory analysis have disappointed the hopes held for them. This owes both to the exaggeration of those hopes and to naive implementation. Too many analysts assumed that, having thrown enough data into a machine, that machine--programmed to perform principal components--would think for them. Analytic sensitivity, creativity, and insight are essential to best application of exploratory analysis.

Adopted in this vein, exploratory analysis, while no panacea, might prove a useful tool for AID. It has greatest potential in areas with good measures of program success, with good data, where a variety of approaches have been tried in a variety of settings, and where intelligent interpretation ensues. It may address directly a primary bugaboo of AID evaluation: the problem of generalizing findings across settings. Exploratory analysis may thus be used to identify combinations of background variables that may render programs more or less successful.

Planned Variation

Related to exploratory data analysis is the plotting of experimental approaches to discover effective program strategies. In its most ornate form, this may borrow from experimental theory intricate designs (factorials, multi-level blocks (with or without confounding), Latin and Greco-Latin Squares) to ensure that all important interactions are tested. Informally, it may be effective simply to survey a diverse range of projects and to identify the more successful as patterns for replication. The latter approach, although hampered by contextual peculiarities, should and does take place within AID.

Planned variation has been most extensively applied in the social realm to education³⁶—a field which ostensibly boasts the experimental prerequisite of measurable goals. The theoretical appeal has, however, been dimmed by practical problems. Project heads are more interested in project success than in faithfully following an experimental protocol. If they think that deviation from protocol will achieve improvement, they are tempted. Effective overall monitoring—which is expensive and rare—is thus required. Informal planned variation may degenerate into a shameless quest for bright spots to justify budget submissions--an attempt to obscure general failure by spotlighting local successes.³⁷

These difficulties have been experienced in AID and teach that planned variation, which may help to identify effective project models, should be invoked with caution. It has greatest prospective value where measures of project success are unambiguous and where a new program departure suggests the need for experimentation to find the best means of achieving program ends.

Modeling

In appropriate doses, various types of modeling—econometric, input-output, network, simulation—have proven value. Successes include John Meyer's model of the transportation system in Colombia, several epidemiological models described by Norman Bailey,³⁸ input-output models for

³⁶ Being the title of an experimental program sequel to Head Start.

³⁷ An example is American Institutes of Research; "It Works"; Palo Alto; 1968--a culled series of exemplary projects within the ESZA Title I Program.

³⁸ The Maximal Theory of Infectious Diseases and its Applications; London: Griffin; 1975.

economic development, and water resource modeling.³⁹ Unfortunately, too many modelers have been seduced by modern computational power to build unmanageably complex models with lower precision than smaller, simpler models. Examples of overextended models in the United States are the massive simulations of the Delaware Estuary and of the Northeast Corridor Transportation System. Culminating the modeling efforts have been the global models of Forrester and Meadows and of Masarovic and Pestel. While many models have technological emphasis, socio-economic modeling is now becoming more common.⁴⁰

Issue-oriented studies are unlikely to be technically furthered by complex computational models. Both simple computational models and complex descriptive models may, however, be useful for exposition—for instance, showing the hypothetical impacts of such diffuse social processes as the erosion of the family unit. Among the most successful of models in this vein are those used illustratively as nodes of interactive communication and of mediation among multiple decision makers. Non-technical persons are involved in constructing and in using the models—processes that lead to faith in the models and to commitment to their prescriptions. Examples are the various ecological impact models of C.S. Holling, the sociological gaming models of Jose Villegas, and the multi-attribute decision models of Ralph Keeney et al.

Modes of Valuation

When projects have variegated effects—each bearing on their overall success—accurate evaluation is complicated. The complexity is reduced

³⁹ See G.E. Toebes and A.R. Rao; "Environmental Research and Development Related to the Hydrosphere"; Paris: Unesco; 1974.

⁴⁰ E.g., P. Flaissner; "A Simulation Model of the Austrian Health Care System"; in N.T.J. Bailey and M. Thompson (eds.); Systems Aspects of Health Planning; Amsterdam: North Holland; 1975

if all components of effect may be calibrated on a single scale. The World Bank has thus greatly simplified and clarified its evaluations by adopting as a universal standard the internal rate of return for projects. AID, which is concerned with a more diverse range of consequences, cannot similarly simplify its own evaluations. Benefit-cost analysis suggests itself as a holistic approach that might standardize and unify project assessments. Unfortunately, it seems that the state of the art at present is not equal to the challenges posed by AID context. Just a few of the methodological problems faced include:

- i. distributional equity—valuing appropriately and differently gains or losses accruing to different groups;
- ii. divergent valuation—reconciling the discrepant weights attached to effects by the U.S. and by different groups within the host nation;
- iii. valuing commodities in a planned economy—having, without the usual useful tool of market valuation, to judge the worth of resources. Second-best paradoxes abound—requiring shadow prices and opportunity costs (in particular, for scarce foreign exchange) to be estimated.
- iv. time discounting—taking into account that developing nations tend to have high social rates of time preference;
- v. secondary and tertiary effects—tracing through the transferred gains and costs to ultimate beneficiaries and losers (including externality effects); and
- vi. institutional effects—valuing impacts that do not accrue as monetary gains to individuals, nor even in a surrogate sense as consumer surplus. Examples are strengthening institutional infrastructure or fostering the autonomy of population subgroups.

These problems will not be soon or easily resolved but should be addressed. Such attempts will help to clarify the underlying philosophical problems of valuation and of multiple effect trade-offs. Complete single-scale reduction of effects—e.g., in a benefit-cost ratio—will not be possible and too zealous efforts to achieve such reduction will be pernicious. Instead, benefit-cost analysis should estimate outcomes along a small number of dimensions—allowing decision-makers to place suitable values on those dimensions. The analyst himself cannot with assurance make these judgments (e.g., how to weight benefits to the hill peoples versus those of the valley—determinations, moreover, likely to vary over time).

Innovative modes of presentation should be essayed. All numbers, vectors, ratios that help decision makers to grasp the essence of their choices have value.⁴¹ In the health field, a recent useful creation is the development of cost-effectiveness analysis.⁴² This enables analysts to avoid the impossible problem of assigning a value to a human life and to concentrate instead on best ways of extending or enhancing lives with fixed resource constraints. The concept of the quality-adjusted life year has been used to capture the difference between morbid life (e.g., end-stage renal failure sustained by dialysis) and fully healthy life. This perspective may help to assess AID programs relating to health and to food.

⁴¹ Although selective choice of indices may be biased and impair decisions.

⁴² See M. Weinstein and W. Stason; "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices"; New England Journal of Medicine; 296(1977): pp. 716-721.

52

Another new and promising mode of project valuation is offered by multi-attribute decision analysis. This sub-discipline is concerned with valuing different aspects of policy decisions and combining these values to estimate aggregate worth. A well-known pioneering effort in this vein was an analysis to determine the best site for a new airport for Mexico City.

Subjective Methodologies

An important insight of the 1970's has been recognizing the limits of purely objective analysis. In all scientific designs, there is a margin where researcher subjectivity reigns: difficult choices that cannot be wholly resolved on objective grounds, but instead require judgment. Better design shrinks but does not eliminate this leeway. Instead then of scorning all subjectivity in analysis—as has been the case in the past—decision makers are learning how to interpret and to use judgmental work. The only type of analysis to be scorned is that which, to discourage challenge and to magnify its own significance, claims complete objectivity.

The roles played by judgments of individuals in analysis take many forms:

1. data estimation. Applied analysis finds that much of its needed data simply does not exist and moreover may be prohibitively expensive to gather. A way around this obstacle is to secure expert judgments on vital parameters. There are, however, pitfalls. Experts have in the past been both wrong and biased. Analysis based on such judgments should therefore lay out the strategy for obtaining them and the qualifications of individuals.

In this way, decision makers will be able to judge how far to trust studies incorporating subjective estimations.

- ii. delphi analysis. A more elaborate strategy for harnessing expert judgment is through the structured use of expert groups. This is the delphi approach⁴³ which solicits individual estimations, checks their consistency with related, known data, invites discussion among experts on discrepancies among them, and obtains reestimation by the same panel. Examples of application in health are to estimate parameters of interest in the swine influenza decision⁴⁴ and to gauge the costs of appropriate care for stroke;⁴⁵
- iii. Bayesian methods. Techniques for combining experimental information with prior subjective knowledge are termed Bayesian. The theoretical advantages of this approach are winning increasingly wider acceptance—to the point where Bayesian analysis is now being formally used in the multi-million-dollar National Health Insurance Study. In decision-making forums, considerable resistance to relying on Bayesian arguments remains;
- iv. expert panels. Going beyond the parameter estimation of delphi methods are expert panel considerations of complex problems. Such panels are usually multi-disciplinary in nature and are invited to extend the purview of their discussions as far as they deem appropriate. They may bound the problem, structure

⁴³N.C. Dalkey; "The Delphi Method: An Experimental Study of Group Opinion"; Santa Monica: The Rand Corporation (RM-5888-PR); 1969.

⁴⁴Stephen Schoenbaum et al.; "The Swine-Influenza Decision"; New England Journal of Medicine; 295(1976); pp. 759-765.

⁴⁵Harry Emlat et al.; Estimated Health Benefits and Costs of Post-Onset Care for Stroke; Baltimore: Analytic Services, Inc.; 1973.

it for thought, and conclude by providing recommendations for action. Salient examples of blue-ribbon panel analyses are the Ford Foundation review of energy issues⁴⁶ and the National Heart and Lung Institute assessment of the totally implantable artificial heart;⁴⁷

- v. advocacy analysis. Advocacy has so often in the past been consciously buried in analysis that the suggestion arises to make it explicit, to manage analysis so that best cases are made for dissonant points of view. This is intended to clarify points and premises of disagreement and to help decision makers estimate the range of prevailing uncertainty. Advocacy analyses most often occur without formal planning as study results are challenged by those with different viewpoints.

The methodologies of subjective analysis are still being developed yet offer promise to help to AID i) where data are often difficult to obtain, ii) where a wide spectrum of divergent value structures are relevant to decisions, and iii) where different perceptual ways of structuring problems (based on variant values, backgrounds, or disciplinary perspectives) are possible.

Conclusions

This cursory review of methods possibly useful in issue studies indicates that there is no evaluative technique for all seasons. Choice of the appropriate technique will depend on 1) the importance of decisions

⁴⁶ S.M. Keeny, op. cit.

⁴⁷ H.P. Green; The Totally Implantable Artificial Heart; Springfield, Virginia: National Technical Information Service (BP-240-320); 1973.

partly dependent on the study; ii) budget constraints; iii) time constraints; iv) the nature of available data; v) the existence of accepted measures for project success; and vi) the nature of arguments that will persuade the decision-making forum.

Sensitive compromise between ideal and possible study design is essential. Study planners need to sense just how far existing data sources may be relied on and when they should be augmented with further information collection. Sometimes limited use of questionnaires will be sufficient; at other times, one will fear selectivity of recall or of response and will turn to more direct methods. Considerations of expense will often dictate use of stratified sample surveys rather than complete censuses.

When important sharp decisions depend on research findings, classical experimental design is attractive. Usually it will be found to be excessively expensive—forcing reliance on quasi-experimental design. Whereas best application of the ideal experimental design is a science, best adaptation of quasi-experimental design and of expository modeling is an art. Even the tightest scientific designs require subjective interpretation.

Within AID, the unique evaluative difficulties place a premium on catering to subjective concerns. The absence of widely accepted measures for project success (especially at the goal level) should lead researchers to present and to explain a variety of statistics—selected to provide decision makers with the best grasp of alternatives—within the cost constraints of the study. Innovative means of incorporating subjective judgments and fitting them to the perspectives of decision makers (e.g.,

through expositive modeling, delphi panels, advocacy analysis) should be designed and pursued. Perhaps the decision makers themselves should be invited to participate in structuring studies. Such participatory analysis has in the past had the dual effects of ensuring that no concerns vital to decision makers were excluded and of giving them greater understanding of—and hence confidence in—the analysis. This could help AID to improve host nation involvement in its studies. Decision maker participation may similarly assist in the difficult problems of valuation. Dexterously managed, such studies may themselves be tools for promoting harmony among parties to decisions.

Precisely because best quasi-experimental design in AID will require flexibly accommodating situation characteristics, we must here avoid dogmatism in prescribing methods.