Development Discussion Papers

Program Evaluation Guidelines and Framework for Ghana Education

Conrad Wesley Snyder, Jr.

Development Discussion Paper No. 760
June 2000

Harvard Institute for International Development

HARVARD UNIVERSITY

Program Evaluation Guidelines and Framework for Ghana Education

Conrad Wesley Snyder, Jr.

Abstract

Ghana education basks in a mythical history of quality and efficiency, where an imposed education system produced many exceptional itinerant scholars and professionals. The apparent quality was an artifact of a highly elite, subsidized, small education system with strong external ties for the privileged few. The system did not provide a broad base for development, was not stabilized within the local cultural and institutional context, and was inherently inequitable. Ghana's legacy is a schooling system characterized by low quality, few resources, and ideational and moral corruption. What happened and what can be done are interesting questions with few answers. Evaluation plays a ritual role in Ghana and it has not become a tool for reflection and learning at the organizational level. This paper presents guidelines and a rational framework as part of a training program for Ghana educators from the Ministry of Education. Evaluation can redirect the system, provide clues for remedial and corrective action, and pull the system out of recurrent failure, but only if there is political will and a dramatic change in organizational culture.

Keywords: educational reform, learning organizations ,monitoring and evaluation

JEL Classification: I20

Wes Snyder is Faculty Associate of HIID, Professor at the University of Montana, and field leader for the HIID Performance Monitoring and Evaluation (PME) team in Ghana. The work in Ghana focuses on professional training and the development of a systemic M&E system. The PME project is part of the Quality Improvement of Primary Schools (QUIPS) program of USAID/Ghana.

Program Evaluation Guidelines and Framework for Ghana Education

Conrad Wesley Snyder, Jr.

As the root of the word 'evaluation' indicates, an evaluation is concerned with and influences the values and attitudinal judgments of individuals concerned with a particular program or system under scrutiny. When effectively carried out, evaluation provides a comprehensive perspective on a program to **inform the judgments and decisions** to be made about and within that program. In order to serve this function, the information provided by the evaluative activities must meet many criteria, for example, importance, accuracy, reliability, relevance, credibility, timeliness, cost effectiveness, and clarity. These qualities are the procedural goals of the evaluation process; therefore, guidelines for evaluation activities are the essential performance markers for attaining these performance requirements.

By making the guidelines for the conduct of an evaluation explicit and consistent, the procedures for reaching these goals can be clarified and improved. In particular, these guidelines should enhance the capacity of Ghana evaluations to achieve eight interrelated purposes:

1. To inform decisions that influence policy formation and development.

Most decisions are made by a pluralistic management community, rather than by a lone decision maker.[1] This means that evaluation will usually operate in the context of accommodation (which implies the negotiation among various interest groups) and only rarely in the context of command (where there is a single authority who decides the course of action from the single perspective). Policy emerges from the contrasts of interests among the potential policy makers.[2] The role of evaluation is to inform these interests so that some element of rationality underlies policy decisions.

2. To recommend courses of action or changes in present activities.

Remediation of program efforts to improve operations or to affect conceptual reconstruction of basic premises of the intervention are typical needs in development education programs. Evaluation can provide information that can assist in program or design modification (developmental evaluation). Revisionist recommendations are an important part of the evaluation process. Since most development contexts can be described as 'in flux,' the activity of evaluation will almost automatically lead to revised or new courses of action and operational changes. Evaluation plays an important role in the provision of feedback to the change process.

---

[1] Lee J. Cronbach, *Designing Evaluations of Educational and Social Programs* (San Francisco, CA: Jossey-Bass, 1982).

[2] S. Raizen, and Peter H. Rossi, *Program Evaluation in Education: When? How? To What Ends?* (Washington, D.C.: National Academy Press, 1981).

3. To clarify program intents and reduce informational uncertainties.

As a result of environmental presses, program intents can shift dramatically during the course of implementation. An evaluation can chart these changes so that program managers can ascertain the viability and acceptability of these changes. The evaluation can also help new staff better understand the program in terms of its own historical development. Sometimes a program can only be rationalized when the stream of events which shaped its boundaries are fully described. In the development context, there is an urgency to get involved in many of the needs and activities that surround a program. It's not unusual for a program to lose its original purpose and expend its resources along its new line of attack or expand its purview and spread thin its resource base. Evaluation can provide the perspective to assess these directions and evaluate the intents of the program (goal evaluation).

4. To elucidate and possibly alter attitudes to the program under scrutiny.

Attitudes are an important part of the context of program implementation. Variation in perspective can alter the quality 'seen' in the program and influence the range of resources and opportunities available to the program in the future. Attitudes also provide a window to implementation issues. By understanding the various orientations to the program, we can unravel the important issues from the myriad of development problems with which any intervention must grapple. Participants and nearby observers of the program 'carry' with them a great deal of informal information about the activities of the program. Tapping this resource can lead to greater insight into the program. Attitudes are not necessarily directly linked to individual behavior, but they do tend to predict general behavior. They must be dealt with to accomplish meaningful change and to fully understand the program-in-context.

5. To encourage commitment to and to ensure the context fit of program activities and goals.

A program can hardly be expected to be worthwhile unless the participants are committed to its success and have effectively disengaged from counter-measures that might undermine its effectiveness. Programs that emerge from international aid assistance projects are sometimes (perhaps even frequently) thrust upon the system (not always just from the outside). An evaluation may be required to sort out the interface dynamics of the program with the dynamics of the extant system. International projects can also produce competitive strategies, either within complex projects of one donor or across several donor.[3]. With multiple, fragmented influences contending for acceptance and adoption by the system actors, an evaluation may be required to address the problem of *pluralist accommodation*.[4] No innovation can hope to fit all contexts and be acceptable to all stakeholders. Adaptive evaluation seeks to fit or adapt the program to a local context or situation that may have multiple cultural and political agenda.

---

[3] For example, Joane Nagel, and Conrad Wesley Snyder, Jr., "International Funding of Educational Development: External Agendas and Internal Adaptations: The Case of Liberia, *Comparative Education* Review, 33, No. 1 (1989): 320-342.

[4] Lee J. Cronbach, S.R. Ambron, S.M. Dornbusch, R.D. Hess, R.C. Hornik, D.C. Phillips, D.F. Walker, and S.S. Weiner, *Toward Reform of Program Evaluation* (San Francisco, CA: Jossey-Bass, 1980).

6. To provide insight into the program activities and possible consequences.

Evaluation has been described as involving both description and judgment (Stake's two 'countenances').[5] In the context of development interventions, a program consists of a network of hypotheses.[6] The network reflects the interrelationships of presumed links that exist among program components. Evaluation 'describes' these components and then 'judges' them in terms of the hypothesized links, which in turn lead to some set of consequences. The description must be rich and deep so that linkage diagnosis is possible, both to improve understanding of the program and serve as a basis for remediation, if necessary. The evaluation serves as a road map of the program. By examining the connections, it becomes possible to follow qualitatively the proposed causal paths that programmatically bring about the accomplishment of change.

7. To inform program management about problems and issues confronting them.

One of the major purposes of formative evaluation is to determine the effectiveness of operating practices (e.g., instructional materials or strategies), articulate the problems and issues in implementation, and ascertain ideas and opinions on how to improve the situation. Every program has weaknesses and strengths once implemented. By providing feedback to program management, evaluation can assist in the revision of operational activities and materials. The point is to make education as efficient and effective as possible.

8. To assess the needs which the program addresses or should address.

Program evaluation is the assessment of value. Needs assessment determines the importance of the elimination or amelioration of problems found in the local setting. Obviously a program should attend to some socially significant problem and lead to some significantly beneficial condition for the target group of the program. Needs assessment establishes the worth of the problem on which the program is focused. Given the changing development context, priorities can change quickly and even the nature of a problem can be altered by the many influences and interventions.[7] The assessment of needs must address the multivariate interrelationships among needs and the inherent subjectivity.

Guidelines for an Evaluation

The standards selected for inclusion in this document are intended as **guidelines** rather than as regulations or requirements, a position consistent with that proposed in other statements of standards.[8] Evaluation guidelines are ideals of conduct. They are not intended to

[5] Robert Stake, "The Countenance of Educational Evaluation," *Teachers College Record*, 68 (1967): 523-540..

[6] David Klaus, *Evaluation Plan For the DEIDS and Related Projects* (Washington, D.C.: American Institutes for Research, 1974).

[7] e.g., Conrad W. Snyder, Jr., & Joane Nagel, *The Struggle Continues! World Bank and African Development Bank Investments in Liberian Educational Development, 1972-1985* (McLean, VA: Institute for International Research, 1985) and in terms of strange loops, see Conrad W. Snyder, Jr., *Strange Loops in Education: Problems for Planning and Progress* Development Discussion Paper No. 690 (Cambridge, MA: Harvard Institute for International Development, 1999).

[8] A review of the literature related to the use of evaluation standards and guidelines included critiques of five primary sources for Program Evaluation Standards: (1) the *Standards for Evaluations of Educational Programs, Projects, and Materials* (The Joint Committee on Standards for Educational Evaluation 1981) and (2) the Joint Committee's subsequent *The Program Evaluation Standards* (1994); (3) the *Standards for Program Evaluation* (Evaluation Research Society 1982); (4) the *Model Program Evaluation Standards* (Foster, Williams, Lyons, Ilon, & Baghi 1986); and (5) *Educational Programme Evaluation Guidelines for Botswana* (Snyder, with Williams 1987). Thanks are expressed to Bernadette Heckle, University of Montana, for her assistance in this effort.

prescribe a specific set of procedures or techniques for the evaluation. Rather, the guidelines are intended to enhance the likelihood of utility, feasibility, propriety, and accuracy of evaluation. They touch all aspects of evaluation, from the focusing and negotiating stages to the assessment of impact of evaluative feedback to the education system. Although every guideline is not appropriate in all circumstances, each approximates the ideal practice in carrying out an evaluation.

The guidelines do not emanate from or support any one specific model or approach to program evaluation.[9] They serve equally well for qualitative or quantitative methodologies, decision or value orientations, beginning or established interventions, and internal or external assessments. They should be revised when found incompatible or incomplete with respect to professionally recognized program evaluation approaches or practices. The guidelines do not intentionally exclude any conventional evaluation method from being considered acceptable or worthwhile when judiciously and correctly implemented. Furthermore, it is important that the guidelines be administered flexibly to accommodate rather than to exclude evaluation efforts of demonstrable utility.

These guidelines should provide a common frame of reference for both evaluators and administrators as to the appropriate expectations of an effective program evaluation. Use of these guidelines should sensitize evaluators and administrators to the elements of an evaluation which otherwise might be missed, and thereby contribute to higher quality program evaluations throughout the education system.

Format of a Guideline

Each guideline is specified by a number, name, statement of standard, clarification of use, and suggestions for application.

Example:

I.1 **Evaluator Credibility:** Evaluation personnel should be selected on the basis of competence, experience, trustworthiness, or trained in evaluation and supervised by credible resource personnel to ensure quality, consistency, impartiality, and ethical practice.
Clarification: Clarification provides a brief explanation of the guideline.
Suggestions: Rules-of-thumb for the applications of the guidelines are also provided.

I. GENERAL GUIDELINES

I.1. **Evaluator Credibility:** Evaluation personnel should be selected on the basis of competence, experience, trustworthiness, or trained in evaluation and supervised by credible resource personnel to ensure quality, consistency, impartiality, and ethical practice.

C*larification:*

The formal capabilities, professional experience, and personal characteristics of the evaluator affect to a great extent the quality of the evaluation and correspondingly, the value placed on the evaluation by the users of the information. Thus, either by selection or

[9] Many of these guidelines resulted from a combination of literature review and committee review (in Botswana). Some standards were added as a result of comparison with the 2nd edition of the Joint Committee on Standards for Educational Evaluation, which included an international review panel.

through training and (supervised) experience, evaluators should exhibit technical competence, substantive knowledge, integrity, and other characteristics that will inspire confidence and credibility.

Suggestions:

- Establish explicit requirements and job descriptions for evaluation positions.
- Provide inservice training where competencies do not match position or task requirements.
- Consider the skill and personality requirements of a given evaluation task when assigning personnel.
- Obtain qualified external technical assistance when necessary.
- A team of professionals with different skills and areas of expertise may best carry out evaluation and provide sufficient internal debate to yield a higher quality evaluation process and report.

I.2. **Political and Social Representativeness:** The evaluation must be attentive to the different adversarial/advocacy groups with respect to any of the evaluative issues to ensure the representativeness of the various viewpoints.

Clarification:

The educational development context has many constituencies that may be overlooked in the rush for problem solutions. All viewpoints must be taken into account, particularly because education has such broad impact on the society. Fortunately, Ghana enjoys a large degree of freedom of information. Any evaluation must contribute constructively to the open debate of important issues, respecting both the responsibility that freedom entails and the need to reflect all viewpoints fairly and appropriately.

Suggestions:

- Identify the key issues and the range of viewpoints that accompany the program.
- Remember that information belongs to the informant, so obtain release permission for sensitive data.
- Debate over issues should be carried out as openly as possible, such that all sides to issues understand conflicting and contrasting viewpoints.
- Be sure that all constituencies have an adequate opportunity to participate in and react to evaluation products.

I.3. **Individual Rights:** Conduct all evaluation activities so that the rights, welfare, dignity, and worth of individuals are respected.

Clarification:

Evaluators must consider both the ethical and the pragmatic implications of all interpersonal interactions with those involved in the evaluation. To the extent that an evaluation has either positive or negative impact on the self-esteem or feelings of individuals or groups, the evaluation may be affected in a correspondingly positive or negative way.

Suggestions:

- Attempt to understand social and cultural values of all participants, especially where they are diverse and/or different from yours.
- Ascertain both the fears and expectations the participants have concerning the evaluation.
- Minimize the disruption the evaluation causes in the participants' environment.
- Be aware of local and national policies and regulations, as well as moral requirements and professional ethics.
- When information is sensitive or embarrassing, and contributes relatively little to the evaluation, consider not collecting or utilizing it.
- Don't compromise professional standards to the extent of concealing any vital information, even if it might inadvertently prove detrimental to someone in a position of accountability.

I.4. **Side Effects and Unanticipated Outcomes:** The evaluator should bring suspected side effects to the attention of decision-makers and other relevant audiences.

Clarification:

Almost invariably, a program, especially a new one, has unanticipated effects—both positive and negative in nature. In a given situation, side effects may be relatively trivial or they may be more important than the program goals. It is essential, therefore, that provision be made to identify and assess side effects so that they may be considered in any evaluation judgments or decisions.

The evaluation itself may have significant side effects on participants or other involved persons. Therefore, while evaluations typically focus on the goals of a program or activity, it is also important to consider potential side effects at all phases of program development and evaluation.

Suggestions:

- The evaluator has a special obligation to be sensitive to potentially harmful side effects of the program and/or the evaluation.
- The earlier side effects are identified, the better; that is, the identification and resolution of potential implementation or operational problems during the planning stage can save considerable time, money, and frustration.
- Side effects should be considered for students, teachers, communities, and all other parties who might be affected.

I.5. **Fiscal Responsibility:** Evaluation finances must be handled prudently and ethically. The Evaluator should use sound accountability procedures for the allocation and expenditure of resources.

Clarification:

Programs vary in complexity. There is the temptation to follow many leads that inadvertently increase costs significantly or aspire to conduct the investigation in terms of research ideals, with little regard for cost implications. Balancing the resources and time available with the need to produce accurate and useful information is a difficult

management goal. Because evaluations work in complex environments, costs can mount quickly in the pursuit of a strong evaluation design and analysis.

Suggestions:

- Specify the costs at the outset, particularly the major costs.
- Maintain accurate records so that the finances can be well managed.
- Gain pre-approvals before taking on more unanticipated costs.
- Remain frugal in using resources. Make sure that any expenditure leads to some useful gain in the evaluation worthiness.

I.6. **Metaevaluation:** The evaluation will be judged itself against relevant local or professional standards. The strengths and weaknesses of the evaluation should be transparent to all stakeholders.

Clarification:

Evaluation is difficult to do, and yet the evaluation must be appraised in order to judge the value of its information and values. Noting limitations and problems in carrying out the evaluation can help in the better interpretation of results or estimating the accuracy of the results. All evaluations attract criticisms and debate. It serves no public function for the evaluation to assume a more exalted position than it rightly merits on objective standards. It's unlikely that international evaluations will have sufficient resources for another evaluator to evaluate the evaluation so the responsibility for reporting on the adequacy of the evaluation lies with the original evaluator.

Suggestions:

- The evaluator has a special obligation to report on the limitations of the procedures and methods employed in the evaluation.
- Maintain good records of the evaluation steps, procedures, measuring instrumentation, methods, unexpected outcomes, noted problems and ambiguities, and analysis limitations.
- Note that not all evaluations require metaevaluation. High stakes judgments may be the best candidates for this extra effort and cost.

I.7. **Evaluation Impact:** Evaluations need to encourage the stakeholders to be involved as much as possible so that they understand the results and valuations, and they are more likely to use the findings in program funding or adjustment.

Clarification:

There's no guarantee that evaluations will be used. Many evaluations are required by project funding requirements, and the involved parties want to divert the evaluation from critical issues or ignore the evaluations altogether. Few managers in development work can effectively use evaluations. This makes the communication effort of the evaluator that much more challenging. One of the major outcomes of evaluation can be a new or deeper understanding of the program or project. New perspectives can yield useful adjustments.

Suggestions:

- To the extent possible, demonstrate the usefulness of various evaluative paths to the key stakeholders.
- Be open and frank in discussions. Criticism alone rarely characterizes good intents well. There are always positive aspects and difficulties.
- Report on interim basis to avoid shocks and gauge areas of particular interest and concern.
- Understand the values of the stakeholders. Small gains in difficult and complex situations can be gratifying even if less than hoped for.
- The evaluator should not take on the responsibility of acting on evaluation results.

II. FOCUSING

II.1. **Program Specification:** The purposes, claims, context, characteristics, and significant forms of the program or activity to be addressed in the evaluation should be specified as precisely as possible.

Clarification:

A mutual understanding (between the evaluator, clients, and participants) of all aspects of the evaluation must be achieved. A detailed description of the above factors can be used as a basis for this understanding. Any evaluation is an evaluation of some 'entity,' i.e., a program, project, physical object, such as instructional material or technique and could be of multiple versions or forms of that 'thing.' The form that the object of the evaluation takes, the surroundings in which it exists, and the purposes it is to achieve must all be understood.

Suggestions:

- Get descriptions of the program being evaluated and its social, political, and economic context from all possible sources (concerned parties, documents, observers) and compile them. Check accuracy by direct observation.
- Ask the client to verify the scope of work and the compilation of information.
- Sufficiently examine the program so that the form(s) of the program in varying contexts can be clearly identified.
- Determine likely influences context will have upon the program and the evaluation process itself.
- If changes in the program or evaluation process occur during the evaluation, document them.

II.2. **Audience Identification:** Identify clients, decision-makers, and potential users of the evaluation results and clarify their information needs, expectations, and priorities. Where appropriate, evaluators also should help identify areas of public interest or concern related to the program or activity to be evaluated.

Clarification:

Diverse audiences are always involved in any evaluation. It is important to identify them, and within resource limitations find out and accommodate their information needs. If evaluation results are based on information that is not relevant, important, or comprehensive enough to satisfy the audiences, the results are likely to be ignored or criticized.

Suggestions:

- Use the most apparent and authoritative audience members to guide you to other possible audiences.
- Contact representatives of the various identified audiences and ascertain their priorities and needs.
- Reach an understanding with the client on the relative importance of the potential audiences and use this information, along with other input, to achieve a realistic balance between equity and influence in defining the evaluation purposes.
- Emphasize the major concerns of each targeted audience so that you don't lose utility in generalizations.
- Don't ignore other audiences that become apparent during the course of the evaluation.

II.3. **Evaluation Approach:** The type of evaluation best suited to the program, to the evaluation purpose, and to the information needs should be identified and its objectives made clear; the range of evaluation activities to be undertaken should be specified.

Clarification:

Once the nature of the program, the setting, the circumstances, and the purposes to be served are understood, it is important to consider which of an array of alternative evaluation approaches is most appropriate. For example, does the nature of this program and its effects suggest a qualitative, participatory approach, or would an external, quantitative approach be more appropriate? Equally important considerations are the timeline and resources available for implementing the evaluation. The evaluation approach may be dictated primarily by time and resource constraints.

Suggestions:

- Get a clear understanding of the program, the setting, the information requirements, the perspectives of significant audiences, and the resource constraints prior to deciding on the evaluation approach to be used.
- For most evaluations, qualitative and quantitative methods compliment each other and can be integrated into a more compelling study than either would provide alone.
- Deciding on a general approach helps to give focus to an evaluation. But rigid adherence to any single approach, in its pure form, is *not* a virtue in program evaluation. Most evaluations break out into several major questions that may call for alternative approaches.

II.4. **Evaluation Costs:** When the cost of an evaluation becomes a serious concern of the client or evaluator, an estimate of the cost of the proposed evaluation and, where

appropriate, of alternatives, should be provided. The estimate should be prudent, ethically responsible, and based on sound accounting principles.

Clarification:

During the development of an evaluation design the evaluator may find that she/he has not fully anticipated the costs of meeting information needs. The cost estimates should be defensibly realistic and cover all elements of the proposed evaluation including costing of district and school personnel time 'donated' to the evaluation.

If the client finds the evaluation too costly or if the evaluator finds the resources inadequate, both parties should consider whether the evaluation design has been pared to what is absolutely essential and necessary, or investigate whether or not a smaller scale evaluation could adequately serve the needs of the client.

Suggestions:

- For online school evaluations, the estimate of evaluation costs may be a process of estimating the amount of time required of the evaluator, and an estimate of required school resources, including personnel time.
- The cost breakdown of a proposed evaluation should be categorized by procedures, phases, materials, and human resources, in order to allow for scrutiny and negotiation of specific parts of the evaluation design.

II.5. **Feasibility of Evaluation:** Determine the feasibility of conducting the evaluation, either informally or through formal evaluability assessment. Reach agreement with clients and decision-makers from the outset that the evaluation is likely to produce information of sufficient value, applicability, and potential use to justify its imposition and cost.

Clarification:

All evaluation procedures, i.e., the particular actions taken in the process of collecting and using information about the object being evaluated, should minimize disruption to the educational processes being evaluated while maximizing the information return for the resources expended. If evaluators do not consider the pragmatic effects of the evaluation, the evaluation procedures may be theoretically sound but unworkable.

Suggestions:

- Select procedures in terms of known constraints, such as budget and available participants. Don't choose procedures or techniques that are good in textbook cases without thinking of their use in the actual situation.
- Verify practicality of procedures with clients and participants.
- Weigh accuracy and practicality. Compromise to the extent possible, but consider abandoning the evaluation if it appears results will be invalid or seriously flawed in a way that could mislead.
- Ensure that personnel are sufficiently qualified to perform the planned procedures.

II.6. **Access to Information:** Establish, together with the client, those officials authorized to release evaluation data, the procedures for releasing the data, and restrictions on access to the data and/or results of the evaluation.

Clarification:

'Right-to-know' considerations must always be balanced against privacy and, in some cases, security requirements (either ethical or legal). An evaluator must identify all possible 'right-to-know' audiences, other requirements that might conflict, and then discuss these with the client/stakeholder. The responsibility of the evaluator and the client is to develop and/or implement procedures to disseminate evaluation information where necessary and to protect it where required. Treat all members of right-to know audiences equally.

Suggestions:

- Advocate the public's right-to-know when possible.
- Be aware of and knowledgeable about relevant legalities.
- Reach formal agreement with the client on disposition of primary documents.
- Specify early those officials authorized to release information and the evaluator's rights and responsibilities for disclosure and dissemination.
- Don't give the client (or anyone else) unilateral authority to revise evaluation documents.
- Don't release partial information that serves only one particular interest.

II.7. **Conflict of Interest:** Identify potential conflicts of interest, and take steps to avoid compromising evaluation processes and results.

Clarification:

The various audiences or interest groups that might seek to influence the evaluation toward some particular concern should be identified. These groups often conflict, or have other political agendas. The evaluator might also be considered an 'interest group' if there is the appearance or reality of some 'personal' stake in the evaluation. It is not always necessary to avoid conflict of interest, but it is necessary to plan how to deal with it; the evaluator must ensure that such conflict does not bias the process, findings, or interpretation of the evaluation.

Suggestions:

- As audiences are identified, assess what advantages or disadvantages to them the evaluation might hold.
- Assess the power structure within the audiences and the object of the evaluation.
- Reassure interest groups that their concerns will be considered.

II.8. **Evaluation Management:** Clearly define accountability for technical and financial management of the evaluation through all phases.

Clarification:

To be accountable, evaluators must use funds for the purposes and procedures specified in the contract, maintain compliance with pertinent regulations, and have transactions verified by standard accounting and auditing procedures. Some of these functions may be retained by the client or delegated to a support group within the organization concerned. However, the evaluator should maintain awareness of all aspects and know the extent to which he/she is held responsible for fiscal matters.

Suggestions:

- Maintain adequate and accurate records of money and time.
- Agree with the client on degree of flexibility in all budget areas.
- Make sure all costs are specified in the budget.
- Be aware of legal and ethical guidelines for use of funds.
- Do not change the evaluation procedures, goals, or whatever without considering the fiscal effects.

II.9. **Documentation of Agreement:** Significant agreements reached during the negotiation phase, including purposes, objectives, schedule, obligations and involvements of all parties, as well as policies and procedures regarding access to the data, should be documented. When plans or conditions change, such changes should be documented. Consider the advisability of having these agreements in writing.

Clarification:

Many evaluations are conducted under formal contract, and law probably requires documentation of the above factors. However, even under the best circumstances, the process of writing such agreements will clarify understanding and protect all parties. Ambiguities will often become apparent during the documentation, and later misunderstandings can often be resolved on the basis of a written agreement. In practice, people are reluctant to document changes that they perceive as shortcomings. The conduct of agreements is a difficult part of evaluation. Given the political nature of judgments, even the agreement process is evidence of problems and difficulties. When clarifications or protection is difficult to obtain, there is usually some underlying organizational problem relating to openness and learning.

Suggestions:

- Have an outside party review agreements whenever it seems advisable.
- Negotiate amendments as the evaluation proceeds, and make sure, when possible, they are on paper and not just oral agreements.
- Don't adhere so rigidly to what is in writing that necessary changes are not considered.
- Don't try to put every aspect on paper; a sense of mutual trust and respect is also very important.

II.10. **Acknowledgement of Limitations:** Evaluators should not accept obligations that exceed their professional qualifications and/or the resources available to them.

Clarification:

The professional qualifications (both training and experience) of the evaluator must represent the range and depth of methods required by the evaluation. An evaluator who conducts evaluations or accepts other obligations outside of his/her professional qualifications may face questions regarding his/her credibility. The credibility of the evaluation report may also be questioned, as well as the validity and reliability of the generated data. An evaluator who accepts obligations that exceed the resources available to him/her may be placed in a position of trying to meet all obligations with fewer resources

than necessary or eliminating other agreed upon obligations. Either choice may threaten the credibility of the evaluation and the evaluator.

Suggestions:

- A country may not have many professional evaluators; based on an assessment of the competencies needed to carry out the evaluation and the documented credentials of those likely to carry out the evaluation, provision should be made for training and supervision within the evaluation plan.
- If adequate time and/or resources are not available for the evaluation, suggest alternatives or options to meet the evaluation needs of the client; leave responsibility for decisions on the alternatives or options with the client.
- Sometimes clients ask for the impossible. The evaluability of the entity must be determined. There are many occasions and situations where evaluation may seem useful and it isn't. Evaluation will only be useful in those circumstances where trust and openness prevail.

III. DESIGNING

III.1. **Specification of Approach:** For all types of evaluations, a clear approach or design should be specified and justified as appropriate to the types of conclusions and inferences to be drawn.

Clarification:

The procedures of an evaluation should be chosen with careful consideration as to the purpose of the evaluation. A description of both procedures and purposes is necessary at the end of the evaluation, so findings and recommendations will be applied and interpreted appropriately.

Suggestions:

- Reach a clear understanding with the client about the purpose(s) of the evaluation, i.e., the uses to which the information will be put.
- Reach agreement with the client on the evaluation design.
- Document the actual implementation of procedures as the evaluation proceeds.
- Communicate with the client regarding major changes of purpose or procedure, and document these changes.
- Do not assume that either purposes or procedures agreed upon at the beginning of the evaluation will remain constant or will be adhered to.

III.2. **Justification of Sampling:** If sampling is to be used, it should first be determined where purposive sampling is appropriate and where representative sampling is appropriate. The method and the details of the sampling methodology (choice of unit, method of selection, time frame, and so on) should be described and justified based on explicit analysis of the requirements of the evaluation, including generalization where relevant.

Clarification:

Sampling is often more economical than collecting data from the whole population. Random sampling is desirable (but often impractical) when it is necessary to generalize from the sample studied to some larger population. Purposive sampling is used when it is desirable to learn and gain an understanding about certain cases or types of cases. The credibility of the evaluation data, and subsequent analysis, interpretation, reporting, and ability to generalize beyond the case or sample under investigation depends in part on the validity of the sampling methodology in relation to the requirements of the evaluation.

Suggestions:

- Select the unit of analysis, e.g., school, classroom, student, etc., which will provide the most appropriate data to meet the evaluation needs.
- Make explicit the reasons why any particular sampling strategy was chosen, and identify any distortions in the data that might result from the chosen strategy. That is, try to anticipate criticisms that might be made of a particular sampling strategy.
- Liaise with the national statistics office or appropriate national agency when contemplating national surveys.

III.3. **Justification of Measurement and Observational Operations:** The measurement procedures and instruments should be specified, described, and justified. Their reliability and validity should be estimated for the population to be measured.

Clarification:

Any measurement must be considered in light of the validity and reliability of the measurement instrument or process. A valid instrument allows sound inferences to be drawn for the defined context (population, etc.). A reliable instrument gives consistent results. Validity and reliability considerations or estimations should be described, and the choice of instruments or observational procedures justified on the basis of these considerations.

Suggestions:

- Validity should always be defined in the specific context of the evaluation.
- Evaluate multiple outcomes, and use multiple instruments or observational procedures to help provide indications of validity.
- The type of reliability desired (stability over time, between instrument forms, etc.) should be specified with regard to the purpose of the measurement. Report actual reliabilities of your data as well as any published data.
- Realize that, in many evaluations, the best that can be achieved is: (1) minimized unreliability, and (2) careful description upon which others can judge validity and reliability.

III.4. **Obtaining Cooperation:** The necessary cooperation of program staff, affected institutions, and members of the community, as well as those directly involved in the evaluation, should be planned and assurances of cooperation obtained where possible.

Clarification:

The cooperation and/or participation of appropriate program stakeholders will have a direct effect on the extent to which the evaluation design is appropriate, the data collection is successful, and the interpretation is credible. Planning the necessary cooperation of persons and groups associated with the program and evaluation will decrease the likelihood of overlooking or undervaluing a program stakeholder at a critical stage. Obtaining assurances of the cooperating persons or groups will alert all parties as to the nature of the expected cooperation. The process of obtaining assurances will additionally alert the evaluator to the possibility of a lack of cooperation or antagonism by certain parties toward the program and the evaluation.

Suggestions:

- Obtain a list of ministry and program staff, affected institutions, and members of the community who will be needed to design, plan, and carry out the evaluation.
- Review this list with relevant administrators and program staff for validation.
- Be explicit about the evaluation's purpose and timeframe to all parties involved with the study.
- Obtain assurances, oral or written as appropriate, of cooperation and/or participation.

III.5. **Plan for Utilization:** Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

Clarification:

The impact of an evaluation (i.e., the effects it will have upon decisions and actions of its audiences) is always a primary concern. A beneficial impact, in which evaluation findings are used to improve educational programs or products, select more cost-effective methods, or stop waste, is, of course, desired. But evaluators must not assume that a good evaluation will necessarily result in a beneficial impact. An evaluator must also consider taking a role as a change agent, and attempt to ensure that the audiences will take note of and utilize the evaluation results in a constructive fashion. Utilization may not occur for many months or years after the evaluation. Evaluation frequently works indirectly on attitudes and knowledge. Action follows slowly.

Suggestions:

- Demonstrate to audiences how results can be useful to them in their work.
- Involve members of the audiences in all facets of the evaluation, to the extent possible.
- Provide interim reports, as appropriate, and always note to the audiences how these reports can apply to them.
- Within limits of time and resources, plan to help audiences assess, interpret, and apply the final results.
- Don't lose interest in the evaluation as soon as the final results are reported.

III.6. **Sources of Error:** An analysis of potential sources of error should be undertaken, and adequate provisions for quality assurance and control should be established.

Clarification:

Errors in the design, planning, and implementation of an evaluation may threaten the generalizability of the evaluation. The most threatening possible errors should be anticipated so that analysis of potential error sources can provide a basis for quality assurance and control.

Suggestions:

- Design and plan the evaluation to meet the model evaluation standards.
- Critically examine the evaluation design and plan for inadequacies or inappropriateness, by using questions such as: (1) Is the focus adequate and appropriate? (2) Have the major audiences and stakeholders been identified?
- Redesign the evaluation models and plans as necessary to control for potential sources of error.
- Repeat the critique-analysis-redesign process as necessary through each implementation stage of the evaluation.

III.7. **Data Collection Plan:** A data collection plan should be developed in advance of data collection.

Clarification:

The data collection plan provides an overall picture of the data collection preparation and tasks. It is necessary to ensure that all preparations and tasks are carried out in a timely and efficient manner. It should relate data collection to each of the evaluation questions including who will collect the data, how the data will be collected, and the schedule for data collection.

The plan should also be helpful in balancing the need for complete and accurate information with the practical realities of available resources and time constraints. The information derived through planning for data collection will also serve as the information base required for the development of additional data collection instruments and procedures, and will help to refine further the evaluation questions.

Suggestions:

- Choose the data collection strategy or combinations of strategies most appropriate to generate the data that will best answer the research questions.
- Specify the instruments, procedures, and schedule for each strategy that will be used in obtaining the desired data.

III.8. **Review for Harmful Effects:** Activities and procedures that entail adverse effects or risks should be subjected to independent review, possibly by an advisory group knowledgeable in evaluation, and then used only with informed consent by parties affected.

Clarification:

Data collection procedures should be reviewed for explicit or implicit adverse effects or risks to subjects or persons related to the evaluation. Examples of adverse effects or risks range from the utilization of an experimental design in which a *clearly desirable* treatment is unnecessarily withheld from a control group (or a questionable treatment imposed on

marginal students), to an interview form resulting in categories of responses which allow the identification and intimidation of respondents.

Procedures that may entail adverse effects or risks should be subjected to independent review by persons knowledgeable in evaluation and in the program being evaluated. In some cases such a review may result in a clear decision by the evaluator; in other cases, the decision may be shared by ministry administrators and/or program personnel.

Suggestions:

- Review the evaluation design and the plan for the data collection strategies, procedures, and schedule with ministry administrators and/or program personnel to review for potential adverse effects.
- Review the proposed innovative program or practices with school personnel to consider the need for procedures to identify students who might be adversely affected by the new program so that alternatives can be provided.

III.9. **Record Design Changes:** Provision should be made for the detection, reconciliation, and documentation of significant departures from the original design.

Clarification:

Monitoring of the evaluation process should be conducted so that the evaluator and/or evaluation sponsor may detect departures from the plan. 'Departures' may be in the areas of sources or types of data, data collection methods, and the data collection schedule. Detected departures should be reconciled so that departures that are determined to be undesirable are assessed as to their effects on the ability to address the evaluation questions, and what corrective measures can be taken. Departures that are determined to be acceptable/desirable should be demonstrated to be valid and reliable alternatives to their counterparts in the original design. All departures should be documented (as appropriate) as an 'update,' to be included in the final evaluation report.

Suggestions:

- Use the data collection schedule, including sources and types of data, and data collection methods as a checklist for monitoring the data collection.
- Note all departures and review for desirability and potential effects on the evaluation.
- Update the data collection schedule to reflect the departures, acknowledging the desirable departures and revising the schedule as necessary to compensate for undesirable departures.

III.10. **Specification of Interim Communication Approach:** Provision should be made for the communication of interim information about the evaluation. Reports should be useful to the targeted audiences, and the process of feedback/informing should be appropriate (e.g., radio, executive summary reports, meetings, technical reports, policy briefs, video).

Clarification:

An evaluation is carried out in order to inform some group of program stakeholders about what is going on in a program, its progress, and its value to the people involved (and

ultimately the general society). Communication is an important feature of formative evaluation in order to render the information useful and appropriate. The design of an evaluation must therefore specify the nature and process of communication to ensure the setup of appropriate mechanisms and the establishment of reasonable expectations. Good communication is frequently the most important contribution a formative evaluation can make.

Suggestions:

- Remember that written reports are not always the best (and certainly not the only) way to communicate. Look into other ways that are appropriate and acceptable to the audiences of the evaluation.
- Reach an explicit agreement with the client on the mechanisms and process of communicating the activities and findings of the evaluation.
- Devote considerable resources to the communication process. It is fundamental to a useful formative evaluation.

IV. INSTRUMENTATION AND DATA COLLECTION

IV.1. **Ensure Objectivity of Data:** The data collection and preparation procedures should be protected by safeguards so that the findings and reports are not distorted by any bias of the data collectors or other interested parties.

Clarification:

Evaluation results and findings can be biased if the facts upon which they are based have not been impartially collected and selected. Considerations of validity and reliability enter into the objectivity of data, but selection of data used in analyses and the preparation of reports also affects objectivity. Errors of omission or commission can occur, either through deliberate attempts to whitewash, carelessness or ignorance, or audience pressure.

Suggestions:

- Ensure that results include perspectives independent from those persons whose work or product is being evaluated.
- Describe all steps taken to ensure reliability and objectivity.
- Seek out and report on biases or prejudices that may have entered into any data collected.
- Strive to achieve and maintain evaluator independence in reporting.
- Do not surrender any authority for data selection or report editing to the client or an audience.

IV.2. **Security of Data:** Data should be handled and stored so that release to unauthorized persons is prevented and access to data, especially to individual data, is limited to those with a legitimate need to know.

Clarification:

Restricting information from arbitrary or unauthorized review is essential to maintaining the confidentiality and confidence of data sources, the integrity of the evaluation, and often

the utility of the results. Persons supplying sensitive information should be informed of procedures for maintaining confidentiality prior to data collection to increase the probability of obtaining accurate, candid information. Data should be stored so that they are not accessible to unauthorized persons. The method of data storage, e.g., paper, computer file/disk, should be determined prior to data collection, and appropriate access restrictions should be in place before data storage.

Suggestions:

- Negotiate secure storage and handling of the data prior to data collection.
- Notify each person supplying data of the procedures for maintaining confidentiality (and anonymity where the potential for retribution exists) as a normal part of introducing the data collection materials.
- When data are disseminated in any way, utilize group results rather than individual results to protect the anonymity of individual suppliers of data.

IV.3. **Documentation of Database:** Documentation should be maintained of the source, method of selection, circumstances of collection, and processing of each set of data.

Clarification:

Documentation of the sources of data and the circumstances of data collection is crucial in order to establish the authenticity of the data; to allow reexamination of the data or a meta-evaluation; and/or for follow-up investigation related to the original sources of data. Documentation of circumstances of data collection and processing of data preparation are central to establishing the validity and reliability of the data collected, as well as the validity of the data analysis and interpretation.

Suggestions:

- Use the data collection plan and subsequent updates as a basis for documenting the sources and circumstances of data collection.
- Make additional notes following each data collection activity that are explicit enough to reflect possible distortion of the data through circumstances, and for accurately reporting data collection activities.
- Describe the sources of information and the actual quantities of each used in the study.
- Document process and personnel involved in the information gathering, and retain instruments used (to include in a technical appendix in the final report).
- State the sampling procedure used to select the information sources, and an indication of the total available sources (population) of each type.

IV.4. **Protecting Against Data Loss:** Appropriate safeguards should be employed to ensure against loss of data through human error or catastrophic events. Once data are loss, there is little that can be done to re-collect some of the in-depth material again. Frequently, evaluations collect data that no other data sources have. Loss data may result in biased reporting.

Clarification:

Lost data may threaten the validity or credibility of the evaluation, may be expensive or time-consuming to replace, or may contain data that cannot be recaptured. All reasonable

safeguards should be taken to prevent damage or loss of data through such events as theft, heat/fire, electrical or water damage. Each method of data storage, such as paper or computer disks, has its own set of hazards and related appropriate safeguards.

Data additionally may be lost through inadequate coding techniques, ranging from codes that do not represent the full breadth and depth of the data to codes that hinder retrieval of or access to the raw data.

Suggestions:

- Create backup copies of data files with highly restricted access.
- Develop and use a data-coding scheme based more on the breadth and depth of the data, and the mode of storage, and less on preconceived notions of the breadth and depth of the final report.
- Consult with persons knowledgeable in each medium of data storage, e.g., computer disks, about dangers and appropriate safeguards and backups.
- Establish rules about the scoring, handling, and processing of data and train the staff accordingly.

IV.5. **Description of Perspectives Represented in Data Sets:** The sources of information should be described in enough detail so that the adequacy of the information can be assessed.

Clarification:

A variety of information sources are important not only for comprehensiveness, but also for cross-checking as well. However, since an evaluator can never gather all information from all sources, the samples collected as the basis for the evaluation will affect the results to some degree. In order to allow clients and audiences to interpret results and have confidence in the recommendations, the nature of the sources of information used must be made clear.

Suggestions:

- Provide an adequate contextual description so that the rationale for the sampling procedure and the sources of information is clear.
- Justify the process and instruments used to gather information in terms of their comprehensiveness. Deliberate exclusion of varying points of view or different perspectives should be avoided, if possible, or at least justified.
- A formative evaluation can be misleading unless it deals explicitly with the pluralistic perspectives likely to pertain to the focal program. People may be reluctant to share their 'real' feelings and knowledge in a superficial data collection approach. Provision must be made to probe deeply and widely for varying perspectives.

IV.6. **Reviewing Data:** The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.

Clarification:

While information is seldom absolutely complete, it is possible to make the data used in an evaluation accurate. There are many stages during information collection where errors can

occur; each stage should be systematically controlled and reviewed to eliminate as many errors and sources of error as possible. Bad data can result in erroneous conclusions and recommendations. In addition, errors in data that are detected by clients can also render them suspicious of the entire evaluation effort.

Suggestions:

- Remember that even highly competent professionals can make numerical transpositions, typographical errors, etc. Machines also make mistakes. Use accuracy checks in all cases.
- Identify common types of errors and make sure all personnel are trained in their detection and/or avoidance.
- Adopt standard procedures for error checking, data storage, and data retrieval; make sure implementation follows defined procedure.
- Don't assume that directions on test instruments, questionnaires, etc., are always clear or will be followed by respondents.

V. DATA ANALYSIS AND INTERPRETATION

V.1. **Specification of- Appropriate Analysis:** Methods used to analyze data should be appropriate to the type of data and to the conclusions to be drawn; the methods, the reasons for choosing them, and their underlying assumptions or limitations must be described clearly.

Clarification:

All data collected, whether qualitative or quantitative, should be synthesized or analyzed in some fashion. Since there is no set of analytic methods that are 'correct' under all circumstances, evaluators must rationally defend their methodology, underlying assumptions, calculations, and conclusions. In the process of analyzing program information, the relationship between evaluation design and appropriate data analysis techniques should be considered.

Suggestions:

- Choose analyses that are appropriate for each type of data collected and for the purpose for which the results are intended.
- Document all processes, including categorization of data, cross-checking of sources, computer runs, etc.
- Verify interim analyses with original sources and/or a more experienced data analyst to confirm or clarify results.
- Examine the results for internal consistency and reasonableness within the total context of the study (i.e., the program, the evaluation design and procedures, and the data).
- Strive for a reasonable balance of rigor and detail with relevance and timeliness when analyzing data.

V.2. **Significance of Results:** When inferential comparisons are made, an indication should be provided of both statistical significance and practical importance.

Clarification:

Statistical significance deals with the probability that differences found between groups, or pre vs. post differences for the same group, will be found for the larger population represented by these groups. Practical significance refers to the educational value or importance attached to the differences, assuming that differences reflect program effects of the indicated magnitude. The evaluator should consider and understand the implications of both, and convey to the evaluation audience(s) an estimate of the magnitude of a given effect (and the associated implications), as well as the statistical significance.

Suggestions:

- The traditional rule of thumb for assessing statistical significance is that the probability of rejecting the null hypothesis (the hypothesis of no difference between the population means) when there is no difference is .05 (alpha = .05). However, when group sizes are small and/or when the consequences of making a wrong decision are not costly, this probability could be increased to .10 or higher.
- A rule of thumb for practical significance is that the standardized effect size (the mean difference divided by the standard deviation of the combined groups) should be at least .25. A standardized effect size of that magnitude indicates that the average student in the superior program is at approximately the 60th percentile of students in the comparison program. Depending on other factors in the specific situation, this rule could be modified.

V.3. **Causal Interpretation:** Cause-and-effect interpretations should be bolstered not only by reference to program rationales but also by recognition and elimination of plausible rival explanations.

Clarification:

Most evaluations will be called on to generate valid causal interpretations. Potential rival explanations for the results that were not explicitly included by the design should be solicited and identified. Rival explanations should be addressed and eliminated; if they cannot be eliminated, this should be accounted for in the report.

Suggestions:

- Submit the evaluation design and data analysis for review by persons associated with the program, requesting validation of the study's interpretations as well as plausible rival explanations.
- Review forthcoming rival explanations for plausibility, given the nature and context of the program, as well as the nature of the evaluation design, database, and analysis.
- Rival explanations that cannot be dismissed should be written into the final evaluation report.

V.4. **Rationale for Interpretation:** The frames, perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

Clarification:

The purpose of an evaluation is often a value judgment; that is, rating or scaling the program evaluated for usefulness, importance, or general worth. This valuation is based upon the evaluation results and other relevant information, and the perspectives of influential audiences. Value is a complex and controversial issue. It is difficult to determine who will make value judgments within the evaluation and what procedures will be used for making the judgments. The client and the evaluator should consider these issues, and decisions about the approach to be used should be described, documented, and justified whenever the evaluator is involved in the valuation process.

Suggestions:

- Consider all relevant bases for interpretation of the results; e.g., project goals, laws or regulations, social standards, humanitarian ideals, organizational rules, bureaucratic concerns, civic organizational considerations, and the rules for the system.
- Consider who will make the evaluation interpretations; the evaluator, the client, an audience, a regulatory body, or some combination of these.
- Consider all techniques or yardsticks that might be used; advocacy reports, Delphi convergence, evaluator/client consensus, national standards, organization documents, etc.
- Evaluations need not be objective to the point of being devoid of evaluation interpretations; however, the extent of such interpretation for which the evaluator is responsible should be made clear.

VI. REPORTING AND DISSEMINATION

VI.1. **Communicating Information:** The final report of an evaluation is only one consideration in the interaction between the program and the evaluation. Evaluation must communicate with the appropriate audiences throughout the evaluation. Alternative mechanisms for communication can be utilized, because the evaluation serves various purposes and various audiences. Guidelines VI. 2-14 refer primarily to written reports, but the same advice applies to interim communication strategies.

Clarification:

It's too late to wait for a final report to communicate with program personnel and relevant audiences. An evaluation must establish a dialogue with its clients to enable formative and dynamic changes in the program during the course of the evaluation. As indicated in Design Guideline III.10, communication may be the most important contribution of the evaluation.

Suggestions:

- Establish communication links as soon as possible, and provide information as often as feasible to maintain these links.
- Analyze the extant information architecture, even if only informally, and utilize existing flow patterns to assist in program improvement. Evaluations can both improve the quality and quantity of information and serve as a catalyst for the establishment of new program links.

VI.2. **Timeliness of Information and Communication:** Release of information should be timely, so that audiences can best use the information. Discussion of evaluation issues should also take place while the issue is in focus and a central concern. Once decisions are made, it's difficult to go back again, even if prudent.

Clarification:

An evaluation is timely when the information is delivered to each audience at a time when they can best use it. The whole notion of formative evaluation is that the evaluation can contribute to the formation of a program. That means that the information must be available at the appropriate time for the formative activities.

Suggestions:

- Find out which major audiences intend to make decisions for which they want evaluative information and at what times.
- Plan the communication approach so that there are sufficient time and resources to meet the anticipated information needs of the formative process.
- Keep the most critical information needs in mind throughout the work, so that they can be met even if less critical needs cannot.
- Attend to the information flow within the program and its system context and determine how best to provide timely augmentation of information to that flow.

VI.3. **Scope and Content of Report(s):** The evaluation report(s) should describe the program being evaluated and its context, as well as the purpose, procedures, findings, and recommendations of the evaluation. This is so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

Clarification:

Too often, much of an evaluation is lost after the reports have been delivered. The final and interim evaluation reports may be the only source available for answering late-occurring questions, or for guiding further application of the findings. Thus, it is important to be as understandable and comprehensive as possible in any evaluation report. The report, no matter in what medium, should be characterized by:

Clarity
Conciseness
Logical development
Well-defined technical terms
Clear tabular or graphic representations
Relevant examples

Such characteristics increase audience understanding, report credibility, and application of the findings.

Suggestions:

- When possible, supplement written reports with other forms of communication, and always provide interim information (see VI.8).

- Address reports directly to the evaluation purposes.
- Provide sufficient context and examples to give meaning to the evaluation as a whole.

VI.4. **Reporting Findings, Recommendations, and Opinions:** Findings should be reported in a manner that distinguishes between the direct empirical results of the evaluation, and related opinions, judgments, recommendations, and speculation.

Clarification:

The report should clearly distinguish between empirical findings, opinions, judgments, recommendations and speculation so that the reader is aware of what constitutes the evaluation findings. If the evaluator applies the findings of the evaluation to policy, through recommendations, then the circumstances and intent of the evaluator's policy recommendations should be made clear to the audience(s).

Suggestions:

- An evaluation makes judgments about program arguments; it is important to differentiate and characterize the data on which judgments are based.
- Be careful in wording statements to keep clear the quality of the authoritative data that justifies the evaluative claims.
- If policy recommendations are to be made, introduce them with a narrative describing the main data used by the evaluator to make such policy recommendations.

VI.5. **Balanced Reporting:** Present findings clearly, completely, and in a balanced manner. The findings should be organized and stated in language understandable to decision-makers and other audiences. Any recommendations should be clearly and explicitly related to the findings. Balance does not imply the lack of judgment. Judgment remains the hallmark of evaluation. But there may be multiple possible judgments, and an evaluation would be more useful if the range of judgments were presented to the stakeholders.

Clarification:

A balanced evaluation is not necessarily one that generates an equal number of positive and negative findings. It is, however, an evaluation that is sensitive to both strengths and weaknesses, therefore requiring a search for side effects as well as intended outcomes. An evaluation report, even if the goal of the evaluation was to find weaknesses, should delineate both strengths and weaknesses of the program that was evaluated. Such findings should also clearly and explicitly relate to any recommendations that are made, whether with the intent of eliminating weaknesses, using strengths to compensate for weaknesses, or retaining unexpected strengths. Frequently, recommendations are considered as the key to an evaluation report, even when this is not the case.

Suggestions:

- List characteristics of the program and, using as many audience perspectives as apply, classify these factors as strengths, weaknesses, or neutral.
- Whenever possible, generate, assess, and report plausible alternative explanations of the findings.

- Don't ignore possible side effects outside of the particular context of the program being evaluated. Report all foreseen implications at the school and national levels.
- Any questionable results should be identified.

VI.6. **Prioritization of Findings and Conclusions:** Findings and recommendations should be presented in a framework that indicates their relative importance.

Clarification:

Presentation of the findings and recommendations in a manner that indicates their relative importance will guide the attention of the reader and subsequent discussions and actions based on the findings and recommendations. The final report should include a comprehensive, well-structured discussion of results. The executive summary should draw attention immediately to the most important, relevant findings and recommendations.

Suggestions:

- Write the executive summary as if it might be the only product of the evaluation to be read by key decision-makers.
- Include all-important findings and conclusions in a logically ordered, direct statement format.

VI.7. **Reporting of Assumptions:** Important assumptions in all phases of the study should be explicitly acknowledged.

Clarification:

Assumptions that guide the evaluation and/or are held by the evaluators might not be subject to investigation, but will influence the interpretation of data, findings, and recommendations. Consequently, assumptions should be explicitly acknowledged, as reasonable and possible, so that the reader will know the social, scientific, and/or programmatic foundations and premises of the evaluation and may be alerted to inherent limitations or biases of the evaluation. For example, an evaluator may be investigating factors relating to student achievement. For reasons of limited resources or time, the evaluator may choose to focus on teacher-student interaction while assuming the achievement motivation of students to be a constant. The findings of the study might be limited in their explanation of student achievement, and the reader should be aware of the study's operating assumptions.

Suggestions:

- Consider and identify possible factors that may in some way affect the subject of the evaluation, but which the evaluator has chosen not to investigate (explicit assumptions). Acknowledge these assumptions in the report.
- Consider factors or circumstances that other stakeholders might have introduced to the investigation, but which the evaluator may feel irrelevant or unimportant (implicit assumptions), and acknowledge these in the report.

VI.8. **Feedback to Participants:** Persons, groups, agencies, and organizations that have contributed to the evaluation should receive feedback appropriate to their needs. They ‘own’ the data that they provide to the evaluation, and new information derived from the data should be shared with them as a product of their investment.

Clarification:

Interim feedback and dissemination of evaluation findings to participants, audiences and clients, should follow from the plans that were made during early stages of the evaluation. Dissemination plans should also be expanded to include any right-to-know audiences that were identified during the course of the evaluation. Certain right-to-know audiences are generally involved in dissemination efforts, including those who:

Commissioned the evaluation (the client)
Are legally responsible for the program
Funded the program
Supplied substantial amounts of evaluation data
Program personnel
Those whose professional status might be affected

Special effort should be made to reach all right-to-know audiences. This responsibility is also shared with the client; but if conflict arises, the evaluator must be willing to relinquish some of this responsibility, and suggest getting the opinion of an appropriate third party.

Suggestions:

- Have representatives of key audiences suggest what findings and recommendations, in what reporting form, would be of most interest and use to them.
- Check draft reports with representatives of the audiences for clarity and accuracy.

VI.9. **Disclosure of Findings:** Disclosure should follow the legal and the proprietary understandings agreed upon in advance, with the evaluator serving as a proponent for the fullest, most open disclosure appropriate.

Clarification:

Evaluation acts, public pronouncements, and written reports should adhere strictly to a code of directness, openness, and completeness. Certain legal and/or regulatory restrictions may have been imposed at the outset of the evaluation, but the evaluator should strive to achieve full and frank disclosure to the extent possible. Audiences are entitled to reports that present clearly and openly both the evaluator's judgments and recommendations, and the information used to formulate them.

Suggestions:

- Present all relevant points of view, both positive and negative.
- Report key factors that enhance or detract from the evaluation's relation to reality, whether discovered before or during the evaluation; discuss frankly any significant implications for findings and recommendations.
- Try to include judgments and recommendations that represent broad, balanced, informed perspectives.
- Don't confuse full and frank with premature disclosure, causing misinterpretation and confusion.
- Do not overemphasize the limitations of the evaluation; be frank but not falsely modest.

VI.10. **Long-term Storage of Data:** The database and associated documentation should be organized in a manner consistent with accessibility policies and procedures.

Clarification:

The database, as well as the final report, should be carefully prepared prior to dissemination and should not include data and documentation that has been agreed upon to be confidential. The evaluation database and documentation should be categorized so that it may be included in report versions in a manner which provides readers, decision-makers, and the public with pertinent information while maintaining the confidentiality of persons associated with the evaluation.

Suggestions:

- Keep the data base and identification of individual data sources separate from the final report, and maintain their storage in a secure place.
- Report the data in a manner that protects the anonymity of the individual suppliers of data, for example, using group results instead of individual results.

VI.11. **Dissemination for Decision-Making:** Evaluation results should be made available to appropriate users before relevant decisions might be made.

Clarification:

Reports should be timely, that is, delivered to audiences at a time when they can best use the information or apply the recommendations. The potential for application of evaluations that are perceived as too late, or which have caused delays in decisions, is greatly reduced. It is infrequent that an evaluation is done for the sake of the evaluative process, rather than the utility of the product. Thus, increase utility by ensuring the information is there when needed.

Suggestions:

- During planning and at every stage of the evaluation, audiences should be asked if and when they intend to make decisions based on evaluation outcomes.
- Allow sufficient time to meet deadlines and requests for interim reports; plan backwards from due dates.
- Identify alternative courses of action in anticipation of unexpected delays.
- Consider tradeoffs, such as smaller sample size or elimination of lesser goals, to achieve timeliness.
- Always inform the client of any difficulties that may cause delays in reporting.

VI.12. **Dissemination for Clarification, Attitude Change, and/or Decision Justification/Assessment:** Evaluations not only serve to inform decisions, but are also useful for the clarification of problems, the communication of arguments to promote attitude change, and the analysis and assessment of previous decisions.

Clarification:

Quite often the most important use of evaluation is the clarification of the problem set confronted by a program. There may be no specific decisions to be made, but the client may just want information about the varied perspectives on the program of the many

people within and around the program. An external evaluator can sometimes reveal new problems and relationships between problems more efficiently and effectively.

Sometimes a program can be bogged down due to attitudes about the program rather than substantive issues. Evaluation can provide information that may help in attitude change. When credible, evaluation can address directly the values that underlie certain attitudes and their relationships.

Also, in the process of examining a program, previous decisions and their effect are assessed. This information can be useful to program management to indicate needed changes in decision approaches or to highlight particularly effective strategies.

Suggestions:

- Be aware in reports of the multiple purposes and impact of evaluation.
- Use the appropriate medium of communication for the purpose of the evaluation report.

VI.13. **Anticipation of Distortion:** Evaluators must anticipate and prevent misunderstandings, misinterpretations, and misuses of evaluative information.

Clarification:

The responsibility of the evaluator does not end with the submission of the final report. The evaluator should also be accessible to users of the report for clarification and interpretation when necessary, and for rebuttal against misinterpretation and misuse of the information and findings. The extent to which possible misinterpretation and misuse can be anticipated can provide guidance to the evaluators in writing the final report. When appropriate, incorporating interpretations of major stakeholders in the final report can lessen the likelihood of misinterpretation and misuse after the submission of the final report. As the likelihood for misinterpretation decreases, so does the extended responsibility for clarification and rebuttal.

Suggestions:

- Submit a draft of the evaluation report to representatives of major audiences for an interpretive review.
- Clarify the report, addressing, or including other interpretations as appropriate.
- Be available to clients/users of the report to the extent that they use the information for relevant improvements.
- Refer problems to the evaluation advisory committee.

VI.14. **Professional Role:** Evaluators must maintain a clear distinction between their role as an evaluator and any advocacy role they choose to adopt.

Clarification:

An advocacy role is distinguishable from a proactive role an evaluator may assume in order to promote the consideration and use of the evaluation results. If an evaluator should choose to take on an advocacy role in favor of a particular perspective or position, then s/he has a professional obligation to all parties to clearly articulate the chosen advocacy role. An advocacy role may or may not be compatible with the evaluator role, depending on the issues and contextual politics of the evaluation.

Evaluators should be aware of the apparent conflict between advocating certain positions and presenting evaluation results. Evaluators may wish to take advocacy stands, but they should not assume that they possess any special status or competence.

Suggestions:

- Consider whether one's professional actions go beyond the evaluator's role of presentation, clarification, and interpretation of the findings.
- If one's actions favor a particular perspective or position, then one should introduce him/herself as an advocate of that position, rather than as an objective evaluator.

Framework for Evaluation

In the development context, there are many programs and projects to evaluate but few people qualified to carry out evaluations or utilize their results well. The natural question following a review of standards is what implications do these guidelines have for a specific evaluation? How do we actually go about an evaluation? The framework presented here begins as a rational model (i.e., linear, logical, and detailed) that is eventually relaxed into a more creative arrangement to deal with development conditions.[10]

The guidelines were organized from general to more specific planning considerations, such as general guidelines, focusing, and finally, designing the evaluation. Then, the action of the evaluative process is contained in instrumentation and data collection and data analysis and interpretation guidelines. Lastly, we consider the communication aspects, reporting and dissemination for the evaluation findings.

General Guidelines

Evaluator Credibility
Political and Social Representativeness
Individual Rights
Side Effects and Unanticipated Outcomes
Fiscal Responsibility
Metaevaluation
Evaluation Impact

Focusing

Program Specification
Audience Identification
Evaluation Approach
Evaluation Costs
Feasibility of Evaluation

[10] Development is a linked sequence of dynamic higher integrations that continues until the possibilities of a given line are exhausted and a relative stability of maturity is reached. This course of integration is marked by increasing differentiation of explanation. [Paraphrased from Bernard J.F. Lonergan, *Insight: A Study of Human Understanding* (New York: Longmans, 1965)]. Stephan Jay Gould disagrees with the biological complexity model in development and offers the possibility that perceived complexity may be an artifact of initial conditions and limits (e.g., a simplicity floor effect). He also counters that simple forms still predominate, but are ignored by selective characterization of development as increased complexity, and increased complexity as progress. In international development, change is evident (development), but progress is illusive. Lonergan's conceptualization remains appealing for those of us who want to consider the role of increasing organizational and systemic complexity that accompanies the economic development of many countries.

Access to Information
Conflict of Interest
Evaluation Management
Documentation of Agreement
Acknowledgement of Limitations

Designing

Specification of Approach
Justification of Sampling
Justification of Measurement and Observation Operations
Obtaining Cooperation
Plan for Utilization
Sources of Error
Data Collection Plan
Review for Harmful Effects
Record Design Changes
Specification of Interim Communication Approach

Instrumentation and Data Collection

Ensure Objectivity of Data
Security of Data
Documentation of Database
Protecting Against Data Loss
Description of Perspectives Represented in Data Sets
Reviewing Data

Data Analysis and Interpretation

Specification of Appropriate Analysis
Significance of Results
Causal Interpretation
Rationale for Interpretation

Reporting and Dissemination

Communicating Information
Timeliness of Information and Communication
Scope and Content of Report(s)
Reporting Findings, Recommendations, and Opinions
Balanced Reporting
Prioritization of Findings and Conclusions
Reporting of Assumptions
Feedback to Participants
Disclosure of Findings
Long-term Storage of Data
Dissemination for Decision-Making
Dissemination for Clarification, Attitude Change, and/or Decision Justification/Assessment
Anticipation of Distortion
Professional Role

These guidelines refer to the behavior of the evaluation, but not its content. Guidance on what kinds of data to collect and how to organize them depends, to some extent, on the selection of an evaluation model. Here, we offer a simple model that can be used with qualitative or quantitative information.

Think of a program to evaluate. There are two main emphases in evaluation: outcomes and processes. This is a simplification, of course, but most program evaluation will have one or both of these foci.

- **Outcome Evaluation** seeks primarily to answer questions of the form: Is the program worthwhile? Is it better than another program? Has it achieved its goals or intents? Is it worth the resources used and should they be continued? Usually, outcome evaluation refers to continuation and funding.
- **Process Evaluation** seeks primarily to answer questions of the form: How could this program be carried out more effectively? What improvements are necessary? Is the program being implemented? Is there compliance with the program operations? Are the inputs to the program timely and complete? Usually, process evaluation leads to decisions about program change and adjustment.

An evaluation can be applied to a unit, an entire organization, a program that cuts across several units or organizations, or specific projects. National monitoring and evaluation is different in methods from project evaluation or evaluations focused on smaller units, but the general framework is the same. The entity can be regarded as a social system pursuing some goal.

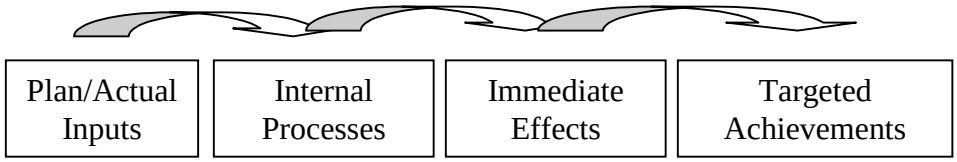
Of course, the long-range goal of educational program evaluation is to assist in the improvement of Ghana's education system. The major premise of evaluation is that better information will lead to better management of the education system, which in turn will lead to better education for the students. The validity of this assertion depends on the efficacy of the many intervening links in the information architecture.

EVALUATION FRAMEWORK[11]

The program or system under evaluation has some ideals or grand intents towards which it works. These intents may be explicit or implicit. They form the criteria for deciding the priority of more specific goals that drive the operations of the program. By definition, the ideals are too ambiguous and general to be measurable and too distant from present realities to be immediately attainable. Programs and projects will have to be completed by a certain time and under more constraints than an ideal situation would call for. Accordingly, we switch attention to specific objectives or *targeted achievements* toward which progress can be assessed.

To achieve these targets, programs must institute and activate certain behaviors or operations. These *internal processes* have *immediate effects* on the clients and stakeholders of the program and on other parts of its environment. The program draws on *planned inputs* from its environment to activate the required (or perceived important) internal operations. In summary then, planned inputs lead to internal operations that produce immediate effects that comprise progress towards targeted achievements, which in turn are the specific objectives intended to move the system towards some ultimate ideals.

[11] The basis for this framework was developed by David Klaus (1974) for application to health programs. It is adapted and elaborated here for application to education programs. The rationale framework serves as a template for the initiation of evaluation but not necessarily its final product.



Rational Evaluation Framework

The model is driven by the ideals. These are the *raisons d'etre* of the program. We track the more observable and immediate components of the program under the assumption that their attainment leads directly to the ultimate attainment of our ideals. We could be wrong, and frequently, time is invested in goal evaluation to be sure that our ideals are good ones.

USE OF THE FRAMEWORK

To use this framework, we work backwards from the ultimate ideals to the targeted achievements to the immediate effects to the internal operations to the planned inputs. Each step constitutes our estimate of what it takes to accomplish the next step. Unless these steps are realistic and measurable, we shall have no way of tracking our progress. This is a management tool so we have to make some compromises to the conditions and resources we have available. Over time, we may be able to improve the approaches and processes we use to assess our accomplishments. If we don't proceed in this explicit fashion, we leave our understandings, impressions, and attitudes to the vicissitudes of poor quality information.

Ultimate Ideals for the System

There are many techniques we can use to find out what we hope to accomplish in the larger system through this program. Usually, the program has evolved from a specifically perceived need. If this were the case, then we'd have the ideal intentions and the information about status, so that we could identify the gap, which is defined as the need. In the rational model, need is precisely defined by this gap and then we can proceed to directly address the specific need. As we deal with the complexities of the real world, however, we find that needs are not clear. Problems are 'wicked.' Needs and problems are highly interrelated, such that it becomes impossible to delineate a clear and unambiguous view of what exactly is needed to improve a system, and single, focused solutions are not effective and sometimes exacerbate the situation. Wicked problems are differentiated from tame problems in the following way.

TAME PROBLEMS	**WICKED PROBLEMS**
Problems can be exhaustively formulated and associated with specific needs.	No definitive formulation is possible; there is no absolutely 'correct view.' There are many views.
Problems have the potential of closure, such that there is a clear solution in sight.	Problems are interconnected to other problems such that solutions have side effects by virtue of these other relationships. There is no clear, single, solution.

Problems are viewed as straightforward with clear root causes.	Problems are complicated with no clearly identifiable single root cause. Every problem can be considered as a symptom of another problem.
Problems exist in a clear, unchanging environment.	Problems exist in a dynamic, largely uncertain environment; the system must be flexible and responsive.
Solutions change the system for the 'better.'	Need to trade off 'goods' and 'bads' even within the same value system; how things work out depends on uncertain interactions among powerful interests.

With wicked problems, there are many explanations for problems. The need(s), which addresses the problem(s), takes on a different form depending on the explanation chosen.

There are several implications to the recognition of the complexity of our systems.

- Must be broad participation in system assessment and program development.
- Must be based on a very wide spectrum of information, gathered from diverse sources.
- Must be on-going so that we remain adaptive to the changes that inevitably occur.

From the set of ideals and the identification of needs, intervention and reform programs are developed to meet these needs, as portrayed in this *Ode to an Intervention Program*.

A program is an operationalized argument,
With which nearly everyone disagrees.
Crafted through political accommodation,
It tries desperately to please.

Doomed to fall short and disappoint
Throughout its stakeholder realm,
It strives to shape some change,
While it doesn't overwhelm.

Evaluation articulates the arguments,
Noting claims and evidence, more or less.
To see if the idyllic program
Has avoided becoming part of the mess.

When the dust settles over our interventions,
Revealing the programmatic confusions,
We shall find a pocket full of arguments,
But no miraculous solutions.

Targeted Achievements for the Program

Targeted achievements are specific targets of the form, 'what will the program accomplish to what standard by when?' Most conventional planning and goal-setting techniques identify targeted achievements.

In education, most targeted achievements, except those connected with formal tests and examinations, are difficult to measure. Accordingly, we turn to educational indicators that provide some useful, related information from which we can infer the quality of the system

or program. Citing many sources, Greaney and Kellaghan[12] characterize indicators in the following ways.

- Quantifiable.
- Each value applies to one point or period of time.
- References a standard, either norm-referenced (synchronic), self-referenced (diachronic), or criterion-referenced, to judge the indicator.
- Relates to important aspects of the system or program.
- Relates to aspects that can be changed.
- Collected frequently enough to assess changes.
- Allows disaggregations to study special groups and regions.
- References a model that is comprehensive enough to describe the important aspects of the system or program and provides a context for interpretation and planning.

Balancing importance with technical adequacy remains a considerable challenge for monitoring and evaluation specialists. To a large extent, the domination of analyses by reliance on test scores 'indicates' the difficulties in using other kinds of indicators. Even if we accept the adequacy of test scores for certain purposes, tests are narrow gauges of achievement and often are forced, for technical, resource, or timing reasons, to remain narrow samples of rich cognitive domains.

Both the program and the measurement of program progress can produce side effects. These are unanticipated effects that emerge from the activities of the program or the observation of the program. When we intervene in complexly organized systems, like education, it is impossible to anticipate all the implications of alterations and actions. Surely many side effects will remain undetected, but good monitoring and evaluation approaches will attempt to uncover possible effects, communicating often and comprehensively with observers and implementers who are involved in the 'experience' of the program. These less obtrusive monitors will often see changes that are subtle and unexpected. Another source of intelligence on side effects comes out of the development of program rationale. As the program description is improved, new relationships and emergent possibilities are more obvious that before. Attention can then be shifted to these alternative hypotheses to better understand their implications in the program's efficacy.

Much of what constitutes the 'evaluative research' model is embedded in this section of evaluation. The targeted achievements and side effects are the measures or indicators of accomplishment for a program. They reflect the learning in the education system. Accordingly, it's here that we find the POEM conceptual framework and data collection design for the program evaluation. Each aspect of the POEM yields a report product that frames the evaluation.

Purpose

- Reasons to do the study. The concepts deemed important.
- *Product*: Evaluation questions.

[12] Vincent Greaney and Thomas Kellaghan, *Monitoring the Learning Outcomes of Education Systems*. (Washington, D.C.: The World Bank, 1996).

Obtrusiveness

- Potential influence of the evaluator on the results. The evaluation activity, itself, can have an impact on the program.
- *Product*: Argument for credibility of the evaluation.

Ecology

- **H**ypotheses: claims for the program.
- **E**ntities: 'things,' units, or individuals studied. These entities have carrier attributes and represent some population of entities.
- **L**ocal Situations: contextual/environmental variables. The concepts are defined/developed in realistic terms for the context.
- **P**rocedures: operations involved in the inquiry. These operations are also selected from a range of possibilities.
- *Product*: The evaluation study itself, presenting the descriptive and evaluative information for consideration..

Maturation

- Time or phases entailed in the phenomena implicated in the evaluation. Development programs take considerable time to effect changes, so the time sampling is important to the detection of the changes.
- *Product*: Generalizability of the findings over time. Evaluation entails causal, deductive arguments that depend on timing and sequence.

POEM links the data with the conceptual underpinnings for the project or program. It provides the warrant for the connection of evidence of targeted achievements with the claims for the program. The argument is formed as **Evidence⇔POEM⇔Claim(s)**. The sophistication of the design depends upon the resources and time available for the evaluation. Although unseen by readers of the evaluation report, the POEM frames the components of the report so that all aspects of the program and the study are illuminated.

Immediate Effects

In laying out the evaluative plan, once we know the targeted achievements and possible side effects, we then look to the linking needs for the *services to be rendered* by the program and the kind of *compliance* or participation expected by the proposed clients and beneficiaries of the program (students, teachers, schools, etc.). Services are usually counts rather than measures, and compliance designates the rates of participation or engagement by the proposed beneficiaries. This represents *interaction status*.

Internal Processes

Services can only be delivered if the internal processes of the program are setup to deliver those services. These processes constitute the *readiness status* of the program to deliver those services and the *scope and capacity* of the effort (one student?, one school?, school district?, the entire national school system?). The argument of the evaluation is inherently deductive. If the program is ready and demonstrates sufficient capacity, then any effects or outcomes can be more easily attributed to its actions. Otherwise, it's difficult to trace the effects of a single program in the complexities of development.

External Processes (Context)

The program is subject to the exigencies of the context within which it must operate. In development contexts, external processes vary a great deal and many programs are contributing to the same goals simultaneously (and possibly, in very different ways). Sometimes they will be conducive to program activities and sometimes they will not be. The history of the program is important to understand the benefits and disadvantages associated with timing and various influences. We can characterize the context as *complementary* to program activities (that is, other activities going on outside the program complement the activities in the program) and *supportive* of its intents (that is, attitudes and political influences encourage the success of the program and are conducive to both the setup of readiness and the compliance of beneficiaries).

Planned Inputs

Planned inputs are the first elements of the proposed implementation and the last component of planning. It's the only component over which we have full control without the many interferences and demands from program and external influences. We can plan inputs, but we cannot guarantee their availability, the implementation of the activities of the program, or any of its outcomes. Planned inputs are assessed on the basis of their *consistency* with the rationales of the program and their *completeness* in terms of what's needed for the program to be undertaken. Evaluating planned inputs is essentially the evaluation of the planning process that went into the development of the program.

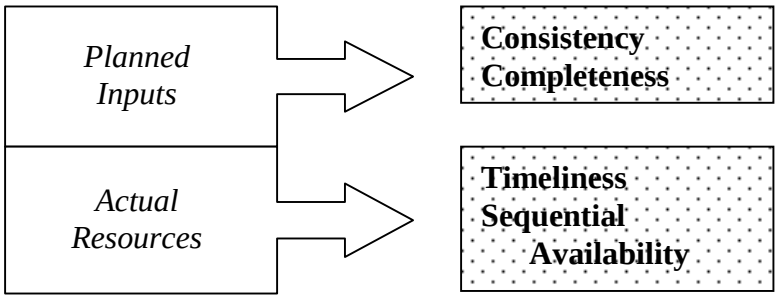
We can begin to develop the outlines of the program rationales at this point. For each of the components, we specify the presumed links and then state the assumptions we are making to justify those links. These assumptions are the first target of our evaluation. If the program rationales aren't consistent, comprehensive, and credible (the 3 C's), then the program hasn't been planned very well. If the assumptions are no longer viable, then the conditions for implementation have changed. If the rationales were never viable, then implementation will not occur, unless conditions change. These assumptions are critical to our understanding of the program. Without understanding, we cannot usefully evaluate the program.

Actual Resources

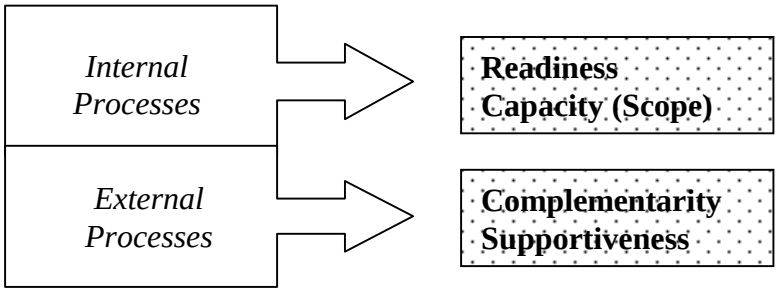
Resource management is an important part of program implementation. Getting the resources at the *right time* and in the *right order* (sequence) is very important. A program generally has some allocation based on budget and/or approved plans. Here we examine the relationship between the plan and the implementation. As we move through the framework, we relate the plan to implementation to effects. Each step is connected by its assumptions, and judged, at least initially, by these designated evaluative criteria.

Components and Evaluative Criteria

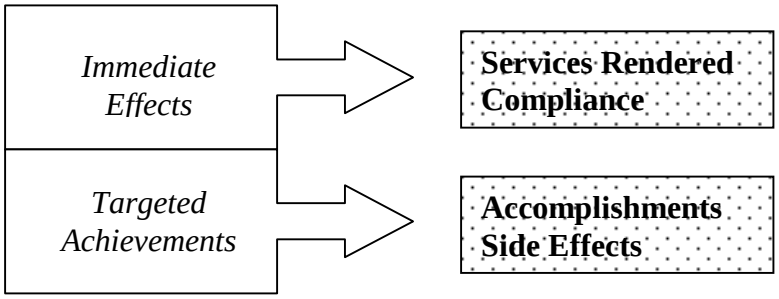
Resources:



Processes:



Claims and Effects:



PROGRAM RATIONALE

How do we link these program components? The object of the specification of the components is to fully characterize the program. Any kind of program analysis method would serve the purpose, as long as it presents a comprehensive picture of the program. Now, in linking the components we need to establish the connection assumptions. These are not formulaic. They will vary by evaluation framework and by particular programmatic considerations. The designation of program rationales is a creative contribution by the evaluator, and it should be a good characterization of the explicit and implicit assumptions of the program. A good program would already have worked out these assumptions, and the evaluator would merely use them or refine them. Usually, because programs are setup

by development professionals who are focused on intervention, documentation of the program is left to the evaluator. Even this step, the characterization of the program, frequently helps people see the larger picture of the organization and its processes. Therefore, it is both useful for the evaluation and immediately useful to open discussions with participants and stakeholders in the program. Recall, at first, no evaluation takes place in these steps. These are descriptive steps that monitor the activities and progress of the program. There's no judgment of value, and no assessment of the assumptions while we are trying to identify them. Description and monitoring alone are valuable. Increasing clarity and understanding are useful management interventions.

Program rationales are the stories of the program. Why does the program need these particular inputs at this time? What processes depend on which inputs? What external conditions enable these processes to operate unimpeded? If enabled and ready, what immediate effects could we expect for each of the processes? Which processes are linked to specific effects? With compliance and cooperation with the program, what targeted achievements can we expect? Are these links tight or loose? Do they rely on external conditions? Are they realistic? When we look at the context of the program, we might find that the rationales of the program are good ones, but the organizational or political climate constrain the results that are implementable or achievable. Under some conditions, neither development intervention nor evaluation work.

Management ignores information in these cases. Only when management is open to information and debate does evaluation contribute usefully to the process. We can go through the motions of evaluation with great technical elegance and sophistication, only to find that there's no user for the evaluation findings. Management may not be capable of mounting any corrective action, if it is pulled and pushed by ambiguous internal influences that mask the larger goals. And if we face an event-based culture, then management is busy addressing apparent concerns within an event, without considering the overall priorities of the organization or the implications of any immediate investments. The relationship between evaluation reports and management has always been loosely coupled. Good management can productively use good evaluation, but good evaluation has little impact if there's no management capacity to deal with the findings.

KINDS OF ARGUMENT

Four kinds of argument form are important for evaluation.[13] Underlying the framework employing program rationales, deductive reasoning is primary. This is elaborated by examples, interviews and data from authorities (or participants who serve as program experts by virtue of their experience with the program), and then reference to implied causality that is couched within the overall deductive argument of the program.

- *Argument by Example*: In describing the program, the evaluator will frequently rely on examples of activities and interactions based on the experiences of the program. These examples highlight or illuminate particular points. They may not be representative, and

[13] Anthony Weston's *A Rulebook for Arguments* (2nd edition, Indianapolis, IN: Hackett Publishing Company, 1992) categorizes five forms of argument: by examples, by analogy, from authority, about causes, and deductive reasoning. We omit analogy because it is rarely used in evaluation studies.

the argument seems strengthened if counterexamples are also offered to illustrate the special contribution of the program. In presenting the examples, background information of the program is also conveyed in an interesting presentation format. The examples provide a vehicle to convey context and draw attention to important points

- *Argument from Authority*: Evaluation gains its data and interpretive power from the many perspectives of the stakeholders and participants of the program. If sources are resistant to release their opinions and experiences, or if they are not well informed, then the foundation of the evaluation is weak. Most surveys in evaluation are purposive and not representative. Their authority underwrites the credibility of the examples used to illuminate the program and the eventual use of these data in causative assessments.
- *Arguments about Causes*: Evaluation draws an arbitrary distinction between *contribution* and *attribution*. In the multiple intervention world of development, attribution is difficult to credit. How do you sort through the many contenders to impact? Contribution, also, is difficult to assess because of the complexities of the context. How do you pick out a single, observable dimension to reflect the changes that have occurred since program initiation? The program rationales are the major tools to logically link causes and effects, but actually mapping program rationales for complicated development interventions is nearly impossible. We end up with many compromises, where causality assessments are made selectively and many connections unexamined.
- *Deductive Reasoning*: The program rationales are structured as 'if A, then B; given A, therefore, B.' This form is called *modus ponens* (the mode of putting), where we put in A, the intervention, and then get the effect, B. Less often, program rationales will take the form of *modus tollens* (the mode of taking), where 'if A, then B; not-B, therefore, not-A.' Here we argue that the absence of effect means the absence of intervention (and we seek implementation deficits to explain the absences). Eventually, we link the program rationales together to form *hypothetical syllogisms*. This form is 'if A, then B; if B, then C; therefore, if A, then C.' We argue that an input leads to some effect, recognizing that the program operated in some appropriate fashion to lead to the effect. Sometimes, by virtue of this logic, evaluation is limited to input audits. The justification is an implied argument of linkages. Input audits are not very satisfying or useful evaluations in the development context, but because of political presses, are often preferred in order to document appropriate use of funds without taking on the more difficult task of sorting out the contributions and attributions of a program.[14] Neither governments nor external agencies want to engage the ambiguities of change. Our development intelligence is unclear on conditions of intervention and self-determination.

CONTEXTUAL AND POLITICAL INFLUENCE

Even as we lay out this rational framework for evaluation, it fits uneasily with experience. Programs don't operate this way, they don't emerge through these logical forms, and they rarely even consider the underlying linkage assumptions in their management. They are

[14] Weston's *disjunctive syllogisms* ('A or B, A; therefore, not-B') and *dilemmas* ('A or B, if A, then C; if B, then D; therefore, C or D') are not used directly in program development, which is linear and simplistic for rapid understanding and funding, and therefore, these forms don't frequently appear in program evaluation.

much more opportunistic and political, if they do anything at all. How do we frame an evaluation on a non-rational program and what are the problems we face to impose it nonetheless? In order to address these questions, the management constraints and powerful political influences have to be considered.

Education programs exist in contexts that are dominated by interests and partisanship, ideologies and beliefs, informational deficits and inadequacies, and institutional myths and cultural influences.[15] This is always a highly politicized environment, partly because of education's ambiguous and multifaceted goals. Unfortunately, this environment diverts attention and, sometimes, efforts away from the primary focus of the educational process—the child. The child has no representation in this dialogue and little influence that matters over the attentions of program developers or actors. The child becomes an abstraction that program dialogue addresses without genuine focus on the real needs and requirements of an actual child. We deal with the interests of the other stakeholders in the system, but ignore the child (particularly again in the rhetoric of program evaluation). Typically in educational programming the emphasis is placed on efficiencies and ideology, rather than creating a better instructional process for a particular child. Evaluation, at its worse, attempts to tighten organizational linkages and exacerbates the confusion of the system.

The cultural context of the program constrains the possibilities for both the program and the evaluation. Ironically, if conditions aren't facilitative for a program, they probably won't be for an evaluation. Both exist only within dynamic, learning organizations that are self-corrective and information oriented. Tom Welsh has distinguished between an event culture and a process culture. In an event culture, an event marks entirely the activities and efforts that take place. There's little connection between one event and another. The event is the focus rather than its underlying intentions, and everything is mobilized for the event as an end in itself. Programs have limited life spans in this kind of culture, and evaluation is superfluous. Within the event there may be apparent progress, but this is quickly dissipated with the end of the event. The event defines the strange loop that temporarily motivates efforts that warp back on themselves without any accumulation of experience or products. Once an event ends, a new defining purpose for action is undertaken with a new event. Only in a process culture is there transfer across activities and programs. Learning can occur within the system and each rendition can better approximate an appropriate contribution to the intended goals. A strange loop prevails unless a process underlies the sets of program activities. Many development contexts reflect an event culture. Evaluation rarely contributes under these circumstances unless it can feed into a value transformation.

The likelihood of evaluation to contribute to the upheaval of an entrenched value system, even if clearly dysfunctional in terms of grander goals, is low. Evaluations rarely occupy the political space necessary to effect immediate or dramatic change. They are sideshows in the accumulation of information, quite often honored in ritual rather than honest inquiry. Their collective hope is for long-term impact based on accumulated evaluation effects. The problem with this hope is that many development contexts are incapable of information

[15] Carol H. Weiss (1995). 'The Four I's of School Reform: How Interests, Ideology, Information, and Institution Affect Teachers and Principals,' *Harvard Education Review*, **65** (4): 571-592.

storage and retrieval, so the evaluation impact is small and transient. Furthermore, there may be no constituency to preserve the information or to retrieve it so there is no way to address this transience. Without demand for information and the capacity to create, store, and retrieve it, there will be no recognition of the strange loops that entrap the system and no creation of new possibilities for real progress. The system remains captured by unique events. In breaking up processes into more limited time spans defined around specific undertakings, there is no reason to link events and activities. Each event justifies the work of individuals without burdening them with grander responsibilities. It is an abdication of responsibility. Welsh has suggested that an individual should take on the specific responsibilities that they can, with the hope that small pockets of responsible individuals will influence the eventual emergence of a new moral order. The 'work ethic' of Western culture served to minimize the captivation of single events. Other cultures identify their ethical base for progress as well or expend their efforts on the treadmill of events.

The problem is further exacerbated by the changes in intent and the increased technical complications of the field. The notion of what's required to manage an education system is changing, and the techniques available to carry out the monitoring and evaluation tasks are improving but much more difficult to use. As Tom Cassidy points out:[16]

> The significant challenge that the shifting knowledge-learning paradigm presents to educational leaders is the need to find ways to help managers, principals, teachers, and students make the shift. An explicit shift to quality, development, and performance demands more of everyone in the system. It is asking everyone to change, first, how they view teaching and learning; second, how they manage and teach; and third, the criteria and methods they use to assess student and school performance. This is a huge shift for many people and system to make.

In a highly politicized environment, where the conceptual and technical demands are increasing, while the context for change is unsupportive, there is little chance that evaluation will have an immediate or dramatic impact. Evaluation is more likely to be overly constrained by the powerful interests in the system. In these constrained environments or those captured in an event culture, which also limits the vision of review and the role of reflection, evaluation can hope only to chip away at political repression and/or provide a larger view to better relate organizationally unlinked activities. Education serves both the child and the society. Because the child is poorly represented in the inherent educational debate (if there is any at all), evaluation must make sure that the child's interests are reflected in the valuing of educational programs. Diversion of activities and resources by powerful interests and event cultures that produce irrelevant and unconnected experiences do not represent the interests of students. Evaluation, at a minimum, can help any system improve its contributions to the students' experiences. Its aspirations and possibilities are limited by its context, like that of the programs that use evaluations for reflection. Strange loops are pernicious and evaluation is a weak directional pointer.

[16] Thomas J. Cassidy, Jr. (1999). *Challenges of Monitoring Education Reform.* Cambridge, MA: Harvard University Graduate School of Education.