2
The CIPP Model for Evaluation

DANIEL L. STUFFLEBEAM
The Evaluation Center, Western Michigan University, MI, USA

This chapter presents the CIPP Evaluation Model, a comprehensive framework for guiding evaluations of programs, projects, personnel, products, institutions, and evaluation systems. This model was developed in the late 1960s to help improve and achieve accountability for U.S. school programs, especially those keyed to improving teaching and learning in urban, inner city school districts. Over the years, the model has been further developed and applied to educational programs both inside and outside the U.S. Also, the model has been adapted and employed in philanthropy, social programs, health professions, business, construction, and the military. It has been employed internally by schools, school districts, universities, charitable foundations, businesses, government agencies, and other organizations; by contracted external evaluators; and by individual teachers, educational administrators, and other professionals desiring to assess and improve their services.[1] This chapter is designed to help educators around the world grasp the model's main concepts, appreciate its wide-ranging applicability, and particularly consider how they can apply it in schools and systems of schools. The model's underlying theme is that evaluation's most important purpose is not to prove, but to improve.

Corresponding to the letters in the acronym CIPP, this model's core concepts are context, input, process, and product evaluation. By employing the four types of evaluation, the evaluator serves several important functions. Context evaluations assess needs, problems, and opportunities within a defined environment; they aid evaluation users to define and assess goals and later reference assessed needs of targeted beneficiaries to judge a school program, course of instruction, counseling service, teacher evaluation system, or other enterprise. Input evaluations assess competing strategies and the work plans and budgets of approaches chosen for implementation; they aid evaluation users to design improvement efforts, develop defensible funding proposals, detail action plans, record the alternative plans that were considered, and record the basis for choosing one approach over the others. Process evaluations monitor, document, and assess activities; they help evaluation users carry out improvement efforts and maintain accountability records of their execution of action plans. Product evaluations

International Handbook of Educational Evaluation, 31–62
T. Kellaghan, D.L. Stufflebeam (eds.)

identify and assess short-term, long-term, intended, and unintended outcomes. They help evaluation users maintain their focus on meeting the needs of students or other beneficiaries; assess and record their level of success in reaching and meeting the beneficiaries' targeted needs; identify intended and unintended side effects; and make informed decisions to continue, stop, or improve the effort.

According to the CIPP Model, evaluations should serve administrators, policy boards, military officers, and other clients; teachers, physicians, counselors, clinicians, engineers, social workers, and other service providers; students, parents, patients, and other beneficiaries; and funding organizations, regulatory bodies, and society at large. Evaluators should present their audiences with evaluations that help develop high quality, needed services and products; help identify and assess alternative improvement options; help assure high quality and ongoing improvement of services; certify the effectiveness of services and products; expose deficient, unneeded, and/or unsafe services and products; and help clarify the factors that influenced an enterprise's success or failure. Thus, the CIPP Model is oriented to administration, development, effective service, prevention of harm, accountability, dissemination, and research.

This chapter introduces the CIPP Model by presenting a general scheme to show relationships among the model's key components. Next, evaluation is defined. The chapter subsequently delineates the CIPP Model's improvement/formative and accountability/summative roles. It follows with a brief discussion of self-evaluation applications of the model. Following discussion of the model's use for improvement purposes, general guidance and an example checklist are provided for using the model for accountability purposes. Context, input, process, and product evaluation are next explained in some detail as applied mainly to group efforts; these explanations include a few cogent examples and a range of relevant techniques. The chapter is concluded with guidelines for designing the four types of evaluation. The Evaluation Center's[2] experiences in applying the model are referenced throughout the chapter.

A GENERAL SCHEMA

Figure 1 portrays the basic elements of the CIPP Model in three concentric circles. The inner circle represents the core values that provide the foundation for one's evaluations. The wheel surrounding the values is divided into four evaluative foci associated with any program or other endeavor: goals, plans, actions, and outcomes. The outer wheel denotes the type of evaluation that serves each of the four evaluative foci. These are context, input, process, and product evaluation.

Each double arrow denotes a two-way relationship between a particular evaluative focus and a type of evaluation. The task of setting goals raises questions for a context evaluation, which in turn provides information for validating or improving goals. Planning improvement efforts generates questions for an input evaluation, which correspondingly provides judgments of plans and

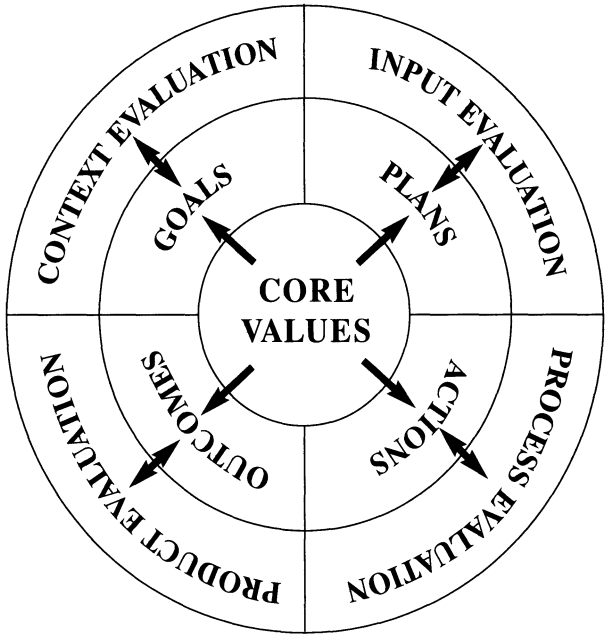

Figure 1: Key Components of the CIPP Evaluation Model and Associated Relationships with Programs

direction for strengthening plans. Improvement activities bring up questions for a process evaluation, which in turn provides judgments of actions and feedback for strengthening them. Accomplishments, lack of accomplishments, and side effects command the attention of product evaluations, which ultimately judge the outcomes and identify needs for achieving better results.

These reciprocal relationships are made functional by grounding evaluations in core values, as denoted by the scheme's inner circle. The root term in evaluation is *value*. This term refers to any of a range of ideals held by a society, group, or individual. Example values – applied in evaluations of U.S. public school programs – are students' meeting of state-defined academic standards, equality of opportunity, human rights, technical excellence, efficient use of resources, safety of products and procedures, and innovative progress. Essentially, evaluators assess the services of an institution, program, or person against a pertinent set of societal, institutional, program, and professional/technical values. The values provide the foundation for deriving the particular evaluative criteria. The criteria, along with questions of stakeholders, lead to clarification of information needs. These, in turn, provide the basis for selecting/constructing the evaluation instruments and procedures and interpreting standards. Evaluators and their clients must regularly employ values clarification as the foundation of their evaluation activities.

A FORMAL DEFINITION OF EVALUATION

The formal definition of evaluation underlying the CIPP Model is as follows:

> Evaluation is the process of delineating, obtaining, providing, and applying descriptive and judgmental information about the merit and worth of some object's goals, design, implementation, and outcomes to guide improvement decisions, provide accountability reports, inform institutionalization/ dissemination decisions, and improve understanding of the involved phenomena.

This definition summarizes the key ideas in the CIPP Model. The definition posits four purposes for evaluation: guiding decisions; providing records for accountability; informing decisions about installing and/or disseminating developed products, programs, and services; and promoting understanding of the dynamics of the examined phenomena. It says the process of evaluation includes four main tasks: delineating, obtaining, providing, and applying information. Hence, trainers should educate evaluators in such areas as systems thinking, group process, decision making, conflict resolution, consensus building, writing reports, communicating findings, and fostering utilization of evaluation results. To fully implement the evaluation process, evaluators also need technical training in collecting, processing, and analyzing information and in developing judgmental conclusions. The definition also notes that evaluators should collect both descriptive and judgmental information; this requires employment of both quantitative and qualitative methods. According to the definition, evaluations should assess goals, designs, implementation, and outcomes, giving rise to the needs, respectively, for context, input, process, and product evaluations. Also highlighted is the fundamental premise that evaluators should invoke the criteria of merit (the evaluand's quality) and worth (its costs and effectiveness in addressing the needs of students or other beneficiaries).

The CIPP Model also posits that evaluators should subject their evaluations and evaluation systems to evaluations and that such metaevaluations should invoke appropriate standards. The standards for judging evaluations that employ the CIPP Model go beyond the traditional standards of internal and external validity employed to judge research studies. The standards employed to judge CIPP evaluations of North American public school programs and personnel include utility, feasibility, propriety, and accuracy (Joint Committee, 1981; 1988; 1994). These standards are targeted to educational evaluations in the U.S. and Canada, but they provide examples that other countries can consider as they develop their own standards for educational evaluations.

THE CIPP MODEL'S IMPROVEMENT/FORMATIVE AND ACCOUNTABILITY/SUMMATIVE ORIENTATIONS

The CIPP Model is designed to serve needs for both formative and summative evaluations. CIPP evaluations are formative when they proactively key the

collection and reporting of information to improvement. They are summative when they look back on completed project or program activities or performances of services, pull together and sum up the value meanings of relevant information, and focus on accountability.

The relationships of improvement/formative and accountability/summative roles of evaluation to context, input, process, and product evaluations are represented in Table 1. This table shows that evaluators may use context, input, process, and product evaluations both to guide development and improvement of programs, projects, or materials – the formative role – and to supply information for accountability – the summative role. Based on this scheme, the evaluator would design and conduct an evaluation to help the responsible teachers, principals, or other service providers plan and carry out a program, project, or service. They would also organize and store pertinent information from the formative evaluation for later use in compiling an accountability/summative evaluation report.

While improvement/formative-oriented information might not answer all the questions of accountability/summative evaluation, it would help answer many of them. In fact, external evaluators who arrive at a program's end often cannot produce an informative accountability/summative evaluation if the project has no evaluative record from the developmental period. A full implementation of the CIPP approach includes documentation of the gathered formative evaluation evidence and how the service providers used it for improvement.

This record helps the external summative evaluator address the following questions:

1. What student or other beneficiary needs were targeted, how pervasive and important were they, how varied were they, how validly were they assessed,

Table 1. The Relevance of Four Evaluation Types to Improvement and Accountability

	Context	*Input*	*Process*	*Product*
Improvement/ Formative orientation	Guidance for choosing goals and assigning priorities	Guidance for choosing a program/ service strategy Input for specifying the procedural design, schedule, and budget	Guidance for implementation	Guidance for termination, continuation, modification, or installation
Accountability/ Summative orientation	Record of goals and priorities and bases for their choice along with a record of assessed needs, opportunities, and problems	Record of chosen strategy and design and reasons for their choice over other alternatives	Record of the actual process and its costs	Record of achievements, assessments compared with needs and costs, and recycling decisions

and did the effort's goals reflect the assessed needs? (addressed by context evaluation)
2. What procedural, staffing, and budgeting plans were adopted to address beneficiaries' needs; how responsive were the plans to the assessed needs; what alternative approaches were considered; in what respects were the selected plans superior to the rejected alternatives; to what extent were the chosen approach and plans feasible, compatible, potentially successful, and cost-effective for meeting beneficiaries' needs? (addressed by input evaluation)
3. To what extent did the staff (or individual service providers) carry out the project plan, how and for what reasons did they have to modify it, and what did the project cost? (addressed by process evaluation)
4. What results – positive and negative as well as intended and unintended – were observed, how did the various stakeholders judge the outcomes' merit and worth, to what extent were the target population's needs met, to what extent were there undesirable side effects, to what extent was the project cost-effective, and to what extent were any poor project outcomes due to inadequate project or service implementation or a faulty design? (addressed by product evaluation)

A CHECKLIST FOR SUMMATIVE EVALUATIONS

As seen in Table 1, applying the CIPP Model proactively to guide decisions yields much of the information needed to complete a retrospective summative evaluation. However, it might omit some important data.

To forestall that possibility, evaluators can apply a checklist designed to cover all variables involved in a comprehensive summative evaluation. An example checklist follows:

1. Overview of the program or particular service (including its boundaries, structure, stakeholders, staff, and resources, and the time frame in which it is examined)
2. Client and audiences for evaluative feedback
3. Program/service background and context
4. Resource/opportunity analysis (service institutions, foundations, staff, volunteers, grant programs, etc.)
5. Targeted/rightful students or other beneficiaries
6. Values, mission, goals, and priorities
7. Planning (process and products)
8. Governance and management (policies and authority/responsibility breakdown)
9. Relationship of the program or service to the surrounding community (services, supporters, detractors, similar programs, etc.)
10. Process (how well was the program or service implemented?)
11. Impact (what classifications and quantities of beneficiaries were reached?)

12. Effectiveness (how well were the beneficiaries served?)
13. Side effects (positive and negative)
14. Costs (e.g., start up and maintenance; personnel, services, and materials; direct and indirect)
15. Sustainability (with and without external funds)
16. Generalizability/transportability (evidence of use and success elsewhere or potential for such use)
17. Comparisons (to alternative program approaches)
18. Significance (e.g., were the outcomes profound and cost-effective?)
19. Recommendations (e.g., needed improvements or continuation versus termination)
20. Reports (tailored to the needs of different audiences)
21. Metaevaluation (did the evaluation meet requirements for utility, propriety, feasibility, and accuracy?)

SELF-EVALUATION APPLICATIONS OF THE CIPP MODEL

It is emphasized that the evaluator need not be an independent evaluator. Often the evaluator appropriately is the teacher, administrator, or other professional who conducts a self-evaluation to improve and be accountable for his/her own services. Consider, for example, how an elementary school teacher might conduct and use formative context, input, process, and product evaluation in relationship to a particular special education student, then compile an accountability/summative evaluation for presentation to parents, administrators, and other parties.

This teacher might conduct a specific context evaluation to tailor instructional goals to the assessed needs of the particular student. After meeting the student, the teacher might review the student's school records, meet with the student's parents, discuss the student's needs and past records of achievement with the student's previous teachers, and engage a school psychologist to conduct a diagnostic evaluation of the student's special needs. Using the obtained information, the teacher would then define the particular learning and developmental goals to be sought for this student during the subsequent school year.

Grounded in these goals, the teacher could next conduct an input evaluation to help chart an appropriate individual educational plan (IEP) for the student. The teacher might begin by obtaining and reviewing IEPs successfully used with students having needs similar to those of this student. Such plans might be obtained from other teachers, the school's instructional resources center, a university's special education department, a government resource center, the teacher's past plans for similar students, etc. The teacher would then screen the identified IEPs to identify those most responsive to the student's diagnosed needs. He or she might next engage the previously involved school psychologist and/or a special education expert to rank the screened IEPs for their potential effectiveness in serving the particular student. The teacher could choose one of

the IEPs to serve as the basis for more specific planning or could merge the most appropriate elements from several plans into a hybrid plan. Next the teacher would add detail to the plan in terms of a schedule and resources and, usually, more specific lesson plans. Subsequently, the teacher could go over the overall plan with the student's parents and the school psychologist and/or special education expert. These exchanges would serve to inform the parents about the draft plan and to obtain their input and that of the school psychologist and/or special education expert for finalizing the plan.

Next, the teacher would conduct a process evaluation in the course of putting the IEP into action. The aim here is to assure that the IEP actually gets implemented and periodically adjusted as needed rather than being set aside and forgotten. The teacher could maintain a dated log of the respective activities of the student, the parents, and the teacher in carrying out the action plan. Periodically, the teacher could meet with the parents to review their child's progress. The teacher would use information from such steps to keep the instructional and home activity process on track, to modify the instructional plan as needed, and to maintain an accountability record of the actual classroom instruction and home support processes.

Throughout the instructional period, the teacher would also conduct a product evaluation. The main purpose would be to assess the extent to which the instruction and learning goals are being achieved and the student's needs met. The teacher would obtain and assess the student's homework products, classroom participation and products, and test results. He or she could also ask the school's psychologist to administer appropriate tests to determine whether the student is overcoming previously assessed deficiencies and whether new needs and problems have surfaced. Also, the teacher periodically could ask the student's parents to report and give their judgments of the student's educational progress. Periodic discussions of such product evaluation information with the parents and school psychologist and/or special education expert would be useful in deciding whether the instructional goals should be modified and how the guiding instructional plan should be strengthened.

Near the end of each marking period and the school year, the teacher could compile all relevant context, input, process, and product evaluation information for this student and write an overall summative evaluation report. Such reports could be much more useful to the student's parents, the school's principal, and subsequent teachers than the simple sets of letter grades.

This example is focused on the most basic elements of educational evaluation – the teacher and a student. After reviewing this illustration, my wife – a former elementary school teacher – said it basically characterizes what good teachers already do. Despite her possible opinion to the contrary, my purpose in including this example is not to belabor classroom practices that are widespread and obvious but to show how the CIPP Model is designed to fit within and support an excellent process of teaching and learning.

The basic notions in this simple illustration can be extrapolated to CIPP evaluations at the classroom, school, and school system levels. In the remainder

of this chapter, the discussion and examples are focused mainly on group rather than individual applications of the CIPP Model.

AN ELABORATION OF THE CIPP CATEGORIES

The matrix in Table 2 is presented as a convenient overview of the essential meanings of context, input, process, and product evaluation. These four types of evaluation are defined in the matrix according to their objectives, methods, and uses. This section also presents certain techniques that evaluators have found useful for conducting each type of evaluation. No one evaluation would likely use all of the referenced techniques. They are presented to give the reader an idea of the range of qualitative and quantitative methods that are potentially applicable in CIPP evaluations.

Context Evaluation

A context evaluation's primary orientation is to identify a target group's needs and thereby provide the criteria for setting goals and judging outcomes. A context evaluation's main contributions are to:

- define a target group of beneficiaries
- identify the group's needs for education or other services
- identify barriers to meeting the assessed needs
- identify resources that could be called upon to help meet the needs
- provide a basis for setting improvement-oriented goals
- provide a basis for judging outcomes of a targeted improvement/service effort

Whatever the target group, administrators and staff can use a context evaluation to set defensible goals and priorities or confirm that present goals and priorities are sound. The context evaluation information also provides the essential criteria for judging an intervention's success. For example, a school's staff may use scores from a diagnostic reading test to later judge whether a reading improvement project corrected the previously identified reading deficiencies of a targeted group of students. As another example, a community health organization might use statistics on the incidence of influenza among a targeted group of senior citizens to assess whether a program of administering flu shots in area supermarkets helped lower the incidence of influenza among these seniors. In these examples, the context information on reading proficiency and influenza incidence provided the baseline information for judging postintervention measures.

Context evaluations may be initiated before, during, or even after a project, course, classroom session, or other enterprise. In the before case, institutions may carry them out as discrete studies to help set goals and priorities. When started during or after a project or other enterprise, institutions will often conduct and

Table 2. Four Types of Evaluation

	Objective	*Method*	*Relation to Decision Making in the Improvement Process*
Context Evaluation	To identify the target population and assess their needs, diagnose *barriers* to meeting the *needs*, identify resources for addressing the needs, judge whether goals and priorities sufficiently reflect the assessed needs, and provide needs-based criteria for judging outcomes	By using such methods as system analysis; diagnostic tests; checklists; secondary data analysis; surveys; document review; literature review; hearings; problem-focused conferences; town meetings; interviews; focus groups; the Delphi technique; school/institution profiles; expert panel site visits; advisory groups; and institutional, program, or service databases	For determining and documenting the *setting* to be served; the *target group* of beneficiaries; the *goals* for improvement; the *priorities* for budgeting time and resources; and the *criteria* for judging outcomes
Input Evaluation	To identify and assess *system capabilities*, alternative program *strategies*, the procedural *design* for implementing the chosen strategy, the staffing plan, the schedule, and the budget, and to document the case for pursuing a particular course of action	By using such methods as literature search, visits to exemplary programs, expert consultants, advocate teams, panel review, and pilot trials to inventory and assess available human and material resources and solution strategies and assess the work plan for relevance, feasibility, cost, and economy	For determining and documenting *sources of support, a* solution *strategy*, a procedural *design*, a staffing plan, a schedule, and a budget, i.e., for *structuring* change activities and providing a basis for judging both the chosen course of action and its implementation
Process Evaluation	To identify or predict *defects* in the work plan or its implementation, to provide feedback for managing the process, and to record and judge the actual work effort	By using such methods as participant observers, independent observers, interviews, document review, and periodic exchange of information with project leaders and staff in order to monitor and provide feedback on the process and record the actual process	For *implementing and refining the work plan and activities,* i.e., for effecting *process control*, and for providing a record of the actual process for later use in judging implementation, interpreting outcomes, and informing replications
Product Evaluation	To collect *descriptions* and *judgments* of outcomes; to relate them to goals and to context, input, and process information; and to interpret their *merit* and *worth*	By measuring intended and unintended outcomes, by collecting judgments of outcomes from stakeholders, by performing both qualitative and quantitative analyses, by comparing outcomes to assessed needs, and by synthesizing findings to reach bottom line conclusions	For deciding to *continue, terminate, modify,* or *refocus* a change activity; and for presenting a clear record of effects (intended and unintended, positive and negative), compared with assessed needs and goals and for interpreting outcomes

report context evaluations in combination with input, process, and product evaluations. Here context evaluations are useful for judging already established goals and for helping the audience assess the effort's success in meeting the assessed needs of the targeted beneficiaries.

The methodology of a context evaluation may involve a variety of measurements of students or members of another target population and their surrounding environment. A usual starting point is to ask the clients and other stakeholders to help define boundaries for the study. Subsequently, evaluators may employ selected techniques to generate hypotheses about needed services or changes in existing services. The techniques might include reviewing documents; analyzing demographic and performance data; conducting hearings and community forums; and interviewing stakeholders.

The evaluators might administer special diagnostic tests to members of the target population. The evaluators might construct a survey instrument to investigate identified hypotheses. Then they could administer the instrument to a carefully defined sample of stakeholders and also make it more generally available to anyone who wishes to provide input. The two sets of responses should be analyzed separately.

The evaluators should also examine existing records to identify performance patterns and background information on the target population. These might include records of involvements of the parents in the education of a targeted group of students, attendance records, school grades, test scores, enrollment in different levels of courses, graduation rates, honors, health histories, immunization records, housing situations, and/or notations by teachers.

Throughout the context evaluation, the evaluators might involve a representative review panel to help clarify the evaluative questions and interpret the findings. They might conduct a meeting – such as a parent-teacher conference or a town meeting – to engage experts and constituents in studying and interpreting the findings and making recommendations. They might also engage focus groups to review the gathered information. The evaluators might use a consensus-building technique to solidify agreements about priority needs and objectives.

After the initial context evaluation, the institution might need to continue collecting, organizing, filing, and reporting context evaluation data. The evaluators could draw selectively from the same set of methods recommended above. They could help stakeholders maintain current information on beneficiaries' characteristics and achievements in a functional input-process-output information system.

Often audiences need to view the effort within both its present setting and its historical context. Considering the relevant history helps the decision makers avoid past mistakes. Thus, the methodology of context evaluation includes historical analysis and literature review as well as methods aimed at characterizing and understanding current environmental circumstances.

A context evaluation may have many constructive uses. It might provide a means by which a school staff talks with its public to gain a shared conception of the school's strengths and weaknesses, needs, opportunities, and priority problems.

An institution might use it to convince a funding agency that a proposed project is directed to an urgent need or to convince an electorate to pass a tax issue in order to meet students' needs better. The context evaluation might be used to set goals for staff development and/or curriculum revision. A school system could also use context evaluation to select particular schools or target populations for priority or emergency assistance. Of course, a school would often use a context evaluation to help students and their parents or advisers focus their attention on developmental areas requiring more progress. Also, an institution could use a context evaluation to help decide how to make the institution stronger by cutting unneeded or ineffective programs. At the national level a government agency might issue an attention-getting report in order to mobilize the public to support a massive program of reform. A famous example of this is the National Commission on Excellence in Education's (1983) report, *A Nation at Risk*, which spawned new U.S. education reform programs. The preceding discussion illustrates how institutions can employ context evaluations to launch needed improvement efforts.

Another use comes later when an institution needs to assess what it accomplished through an improvement project. Here the institution assesses whether its investment in improvement effectively addressed the targeted needs and goals. The institution also refers to context evaluation findings to assess the relevance of project plans. Also, at the end of a project, context evaluation records are pertinent for defending the project's goals and priorities. Considering these uses, a school or other institution can benefit greatly by grounding improvement efforts in sound context evaluations.

The Program Profile Technique, as Applied in Context Evaluations

As noted above, many methods are useful in conducting context evaluations. Evaluators at the Western Michigan University Evaluation Center have devised an overall approach labeled the *Program Profile Technique*. This technique includes:

- a checklist to collect data from a variety of sources about relevant history; current environment; constituent needs; system problems and opportunities; and program structure, operations, and achievement
- a pertinent database
- periodic reports that characterize the program's background, environmental circumstances, and present status
- feedback workshops to the client and designated stakeholders

Using this technique evaluators can maintain a dynamic baseline of information and employ it to keep their audiences informed about the program's status and environment. The successive profile reports present an evolving picture of beneficiaries' needs, objectives, and external forces, and how these relate to program

design, activities, expenditures, and outcomes. In examining such reports, clients and other interested parties gain a holistic picture of the program's progress within its context.

Analysis of Patient Records, as a Procedure for Context Evaluation in Individual Medical Practice

Context evaluations are needed to guide and assess the performance of individual professionals as well as programs. A technique of use in conducting a context evaluation related to improvement needs of individual physicians is what might be labeled the *Compilation and Analysis of Patient Records* (see Manning & DeBakey, 1987). Many such records are routinely completed and stored as a part of the doctor-patient process, including patient files, hospital charts, and insurance forms. In addition, a physician might maintain a card file on unusual, little understood, or otherwise interesting patient problems. This helps the physician gain a historical perspective on such cases. Patient records are a valuable source of context evaluation information. A doctor can use such records to:

- identify most prevalent patient needs
- identify seasonal patterns in patient problems and needs
- select practice areas for improvement
- select appropriate continuing medical education experiences
- better plan services to patients
- change referral and diagnostic practices

The physician can also compare baseline measures with later measures to evaluate improvement efforts. Context evaluation questions that doctors might answer by analyzing patient records include the following:

- What illnesses and types of accidents are most prevalent among the doctor's patients?
- What are the important systematic variations in illnesses and accidents, aligned with seasons and with the patients' age, gender, and occupation?
- To what extent do the doctor's patients evidence chronic problems that treatments help only temporarily?
- What diagnostic tests and procedures does the doctor use most frequently?
- What are relative levels of usefulness and cost-effectiveness of the diagnostic tests frequently ordered by the doctor?
- What types of problems does the doctor typically treat without referral?
- What types of problems does the doctor typically refer to other professionals?
- What are the success rates, at least relative absence of complaints, of referrals to the different referral services?
- To what extent are patients' records complete, clear, and up to date?

- To what extent are patients' immunizations up to date and inclusive of what they need?
- To what extent have patients been taking physical examinations and other needed tests on an appropriate schedule?
- To what extent do the patient records reflect success in managing weight, blood pressure, and cholesterol?
- To what extent do the doctor's patients take flu shots and with what outcomes?
- What are the numbers and types of complaints from patients and/or other health professionals about the doctor's practice?
- To what extent do the patients pay their bills on time?
- To what extent are the doctor's charges within rates set by third-party payers?

The *Compilation and Analysis of Patient Records* procedure is a valuable means of answering questions, such as those listed above. Individual doctors can use this technique to look for weaknesses and strengths in all aspects of their practice, then formulate improvement goals. Medical educators can also usefully employ the technique in cooperation with doctors to set appropriate goals for individualized continuing medical education services.

This technique fits within a chapter on educational evaluation because it applies to the continuing education of physicians. Probably the technique could be adapted for use in providing evaluative guidance for the continuing education of particular teachers, counselors, administrators, and other educators. Certainly, all such professionals need continuing education targeted to their needs. Also, all of them have records associated with their work – such as instructional plans, budgets, feedback from parents, evaluations by supervisors, and students' test results. Such records are useful for identifying areas of one's professional practice that should be improved.

Input Evaluation

An input evaluation's main orientation is to help prescribe a course of action by which to make needed changes. It does this by searching out and critically examining potentially relevant approaches, including the one(s) already being used. Input evaluations can help client groups choose a "best buy" approach when they search out and assess options. An approach that predictably would exceed the performance of others will have no possibility of impact if a planning group does not identify it, compare its merits to those of critical competitors, and choose it for implementation.

Once an approach has been chosen, an input evaluation next assists educators or other professionals prepare the chosen approach for execution. It should also search the pertinent environment for political barriers, financial or legal constraints, and available resources. An input evaluation's overall intent is to help administrators and staff examine alternative strategies for addressing assessed

needs of targeted beneficiaries and evolve a workable plan. A sound input evaluation also helps clients avoid the wasteful practice of pursuing proposed innovations that predictably would fail or at least waste resources.

Evaluators conduct input evaluations in several stages. These occur in no set sequence. An evaluator might first review the state of practice in meeting the specified needs and objectives. This could include:

- reviewing relevant literature
- visiting exemplary programs
- consulting experts
- querying pertinent information services (including those on the World Wide Web)
- reviewing a pertinent article in *Consumer Reports* or a similar publication that critically reviews available products and services
- inviting proposals from staff or potential contractors

Evaluators would set up a file to facilitate storage and retrieval of the information. They might engage a study group to investigate it. They might conduct a special planning seminar to analyze the material. The evaluators would use the information to locate potentially acceptable solution strategies. They would rate promising approaches on relevant criteria. Example criteria are listed below:

- responsiveness to priority system needs
- potential effectiveness
- fit with existing services
- propriety
- affordability
- political viability
- administrative feasibility

Next the evaluators could advise the clients about whether they should seek a novel solution. In seeking an innovation, the clients and evaluators might document the criteria the innovation should meet, structure a request for proposal, obtain competing proposals, and rate them on the chosen criteria. Subsequently, the evaluators might rank the potentially acceptable proposals and suggest how the client group could combine their best features. The evaluators might conduct a hearing or panel discussion to obtain additional information. They could ask staff, administrators, and potential beneficiaries to react and express any concerns. They would also appraise resources and barriers that should be considered when installing the intervention. The clients could then use the accumulated information to design what they see as the best combination strategy and action plan.

Input evaluations have several applications. A chief one is in preparing a proposal for submission to a funding agency or policy board. Another is to assess one's existing practice, whether or not it seems satisfactory, against what is being done elsewhere and proposed in the literature. Input evaluations have been used

in a number of U.S. school districts to decide whether locally generated proposals for innovation would likely be cost-effective. One school district used an input evaluation to generate and assess alternative architectural designs for new school buildings. The Southwest Regional Educational Laboratory used an input evaluation to help historically antagonistic groups agree on how to use ten million dollars to serve the education needs of migrant children. In addition to informing and facilitating decisions, input evaluation records help authorities defend their choice of one course of action above other possibilities. School administrators and school boards can find input evaluation records useful when they must publicly defend sizable expenditures for new programs.

The Advocacy Teams Technique as Used in Input Evaluations

The Advocacy Teams Technique is a procedure designed specifically for conducting input evaluations. This technique is especially applicable in situations where institutions lack effective means to meet specified needs and where stakeholders hold opposing views on what strategy the institution should adopt. The evaluators convene two or more teams of experts and stakeholders. They give the teams the goals, background data on assessed needs, specifications for a solution strategy, and criteria for evaluating the teams' proposed strategies. The teams may be staffed to match members' preferences and expertise to the nature of the proposed strategies. Evaluators should do so, especially if stakeholders severely disagree about what type of approach they would accept. The advocacy teams then compete, preferably in isolation from each other, to develop a "winning solution strategy." A panel of experts and stakeholders rates the advocacy team reports on the predetermined criteria. The institution might also field-test the teams' proposed strategies. Subsequently, the institution would operationalize the winning strategy. Alternatively, it might combine and operationalize the best features of the two or more competing strategies.

The advocacy teams technique's advantages are that it provides a systematic approach for:

- designing interventions to meet assessed needs
- generating and assessing competing strategies
- exploiting bias and competition in a constructive search for alternatives
- addressing controversy and breaking down stalemates that stand in the way of progress
- involving personnel from the adopting system in devising, assessing, and operationalizing improvement programs
- documenting why a particular solution strategy was selected

Additional information, including a technical manual and the results of five field tests of the technique, is available in a doctoral dissertation by Diane Reinhard (1972).

Process Evaluation

In essence, a process evaluation is an ongoing check on a plan's implementation plus documentation of the process. One objective is to provide staff and managers feedback about the extent to which they are carrying out planned activities on schedule, as planned, and efficiently. Another is to guide staff appropriately to modify and improve the plan. Typically, staffs cannot determine all aspects of a plan when a project starts. Also, they must alter the plan if some initial decisions are unsound or need to be changed due to new conditions. Still another objective is to periodically assess the extent to which participants accept and can carry out their roles. A process evaluation should contrast activities with the plan, describe implementation problems, and assess how well the staff addressed them. It should document and analyze the efforts' costs. Periodically, it should present staff with timely feedback they can use to strengthen their efforts. Finally, it should report how observers and participants judged the process's quality. Also, it provides a detailed record of the actual process of implementation.

The linchpin of a sound process evaluation is the process evaluator. More often than not, a staff's failure to obtain guidance for implementation and to document their activities stems from a failure to assign anyone to do this work. Sponsors and institutions too often assume erroneously that the managers and staff will adequately log and evaluate process as a normal part of their assignments. Staff can routinely do some review and documentation through activities such as staff meetings and minutes of the meetings. However, these activities do not fulfill the requirements of a sound process evaluation. Beyond lacking the time to do adequate process review, analysis, and documentation, staff also lack the important element of an independent perspective. Experience has shown that project staffs can usually meet process evaluation requirements well only by assigning an evaluator to provide ongoing review, feedback, and documentation.

A process evaluator has much work to do in monitoring, documenting, and judging an intervention. The following scenario illustrates what he or she might do. Initially, the process evaluator could review the relevant strategy and work plans and any prior background evaluation to identify what planned activities they should monitor. Possible examples are staff training, project planning, staff collaboration, materials development, budget and expenditures, management of the project library, maintenance of equipment, counseling students, meeting parents, tutoring students, skill or interest grouping of students, classroom instruction, classroom assessment, field trips, homework assignments, analysis and use of standardized test results, use of diagnostic tests, and reporting progress. Beyond looking at the elements of work plans, evaluators might also periodically consult a broadly representative review panel. The evaluator could ask the panelists to identify important concerns and questions that the process evaluation should address. Other questions of relevance will occur to the evaluator in observing activities, examining records and other pertinent documents; providing feedback; and interacting with staff, beneficiaries, and the review panel.

With questions and concerns such as those mentioned above in mind, the process evaluator could develop a general schedule of data collection activities and begin carrying them out. Initially, these probably should be informal and as unobtrusive as possible so as not to threaten staff, get in their way, or constrain or interfere with the process. Subsequently, as rapport develops, the process evaluator can use a more structured approach. At the outset, the process evaluator should get an overview of how the work is going. He or she could visit and observe centers of activity; review pertinent documents (especially the work plans, budgets, expenditure records, and minutes of meetings); attend staff meetings; interview key staff; and interview students, parents, and other beneficiaries. The process evaluator then could prepare a brief report that summarizes the data collection plan, findings, and observed issues. He or she should highlight existing or impending process problems that the staff should address. The evaluator could then report the findings at a staff meeting and invite discussion.

He or she might invite the staff's director to lead a discussion of the report. The project team could then use the report for reflection and planning as they see fit. Also, the process evaluator could review plans for further data collection and subsequent reports with the staff and ask them to react to the plans. Staff members could say what information they would find most useful at future meetings. They could also suggest how the evaluator could best collect certain items of information. These might include observations, staff-kept diaries, interviews, or questionnaires. The evaluator should also ask the staff to say when they could best use subsequent reports. Using this feedback, the evaluator would schedule future feedback sessions. He or she would modify the data collection plan as appropriate and proceed accordingly. The evaluator should continually show that process evaluation helps staff carry out their work through a kind of quality assurance and ongoing problem-solving process. He or she should also sustain the effort to document the actual process and lessons learned.

The evaluator should periodically report on how well the staff carried out the work plan. He or she should describe main deviations from the plan and should point out noteworthy variations concerning how different persons, groups, and/or sites are carrying out the plan. He or she should also characterize and assess the ongoing planning activity.

Staff members use process evaluation to guide activities, correct faulty plans, maintain accountability records, enhance exchange and communication, and foster collaboration. Some managers use regularly scheduled process evaluation feedback sessions to keep staff "on their toes" and abreast of their responsibilities. Process evaluation records are useful for accountability, since funding agencies, policy boards, and constituents typically want objective and substantive confirmation of whether grantees did what they had proposed. Process evaluations can also help external audiences learn what was done in an enterprise in case they want to conduct a similar one. Such information is also useful to new staff, as a part of their orientation to what has gone before. Moreover, process evaluation information is vital for interpreting product evaluation results. One

needs to learn what was done in a project before deciding why program outcomes turned out as they did.

Traveling Observer Technique for Use in Process Evaluations

Over the years, The Evaluation Center has developed and employed a procedure labeled the *Traveling Observer Technique* (Evers, 1980; Reed, 1991; Thompson, 1986). This technique most heavily addresses process evaluation data requirements but, like other techniques, also provides data of use in context, input, and product evaluations. The technique involves sending a preprogrammed investigator into a program's field sites. This evaluator investigates and characterizes how the staffs are carrying out the project at the different sites. He or she reports the findings to the other evaluation team members. This investigator may participate in feedback sessions provided to the client group.

The traveling observer (TO) follows a set schedule of data collection and writes and delivers reports according to preestablished formats and reporting specifications. Before entering the field, the TO develops a traveling observer handbook (Alexander, 1974; Nowakowski, 1974; Reed, 1989; Sandberg, 1986; Sumida, 1994). The TO develops the handbook under the principal evaluator's supervision. They tailor this evaluation tool to the particular evaluation's questions. This handbook includes the following:

- traveling observer's credentials
- evaluation questions
- description of the study sites and program activities
- contact personnel and phone numbers
- maps showing project locations
- data sources suggested, including interviewees and pertinent documents
- protocols for contacting field personnel and obtaining needed permissions and cooperation
- rules concerning professional behavior expected
- safeguards to help the TO avoid cooptation by program staff
- sampling plans, including both preset samples and exploratory grapevine sampling
- recommended data collection procedures
- data collection instruments
- data collection schedule
- daily log/diary format
- rules for processing information and keeping it secure
- the audience for TO feedback
- reporting specifications and schedule, including interim progress reports, briefing sessions, and expense reports
- criteria for judging TO reports

- rules about communicating/disseminating findings, including provisions for reporting to those who supplied data for the TO study
- any responsibilities for scheduling and facilitating follow-up investigations, e.g., by a site visit team of experts
- issues that may arise and what to do about them
- form for the TO's periodic self-assessment
- budget to support the TO work, including spending limitations

In an early application of this technique, The Evaluation Center sent out traveling observers as "advance persons" to do initial investigation on two $5 million statewide National Science Foundation programs. The Center assigned the TOs to prepare the way for follow-up site visits by teams of experts. These teams included national experts in science, mathematics, technology, evaluation, and education. Each program included many projects at many sites across the state. The evaluation budget was insufficient to send the five-member teams of "high priced" experts to all the potentially important sites. Instead, the Center programmed and sent TOs to study the program in each state. Each TO spent two weeks investigating the program and prepared a report. Their reports included a tentative site visit agenda for the follow-up teams of experts. The TOs also contacted program personnel to prepare them for the follow-up visits and gain their understanding and support for the evaluation. On the first day of the team site visits, each TO distributed the TO report and explained the results. The TOs also oriented the teams to the geography, politics, personalities, etc., in the program. They presented the teams with a tentative site visit agenda and answered their questions. The TO's recommended plans for the site visit team included sending different members of the site team to different project sites and some total team meetings with key program personnel. During the week-long team visits, the TOs remained accessible by phone so that they could address the needs of the site visit team. At the end of this study, the Center engaged Michael Scriven to evaluate the evaluation. He reported that the TO reports were so informative that, except for the credibility added by the national experts, the TOs could have successfully evaluated the programs without the experts. Overall, The Evaluation Center has found that the Traveling Observer technique is a powerful evaluation tool; it is systematic, flexible, efficient, and inexpensive.

Product Evaluation

The purpose of a product evaluation is to measure, interpret, and judge an enterprise's achievements. Its main objective is to ascertain the extent to which the evaluand met the needs of all the rightful beneficiaries. Feedback about achievements is important both during an activity cycle and at its conclusion. A product evaluation should assess intended and unintended outcomes and positive and negative outcomes. It should be especially attentive to harmful side

effects. Moreover, evaluators should often extend a product evaluation to assess long-term outcomes.

A product evaluation should gather and analyze judgments of the enterprise by stakeholders and relevant experts. Sometimes it should compare the effort's outcomes with those of similar enterprises. Frequently, the client wants to know whether the enterprise achieved its goals and whether the outcomes were worth the investment. When indicated, evaluators should interpret whether poor implementation of the work plan caused poor outcomes. Finally, a product evaluation should usually view outcomes from several vantage points: in the aggregate, for subgroups, and sometimes for individuals.

Product evaluations follow no set algorithm, but many methods are applicable. Evaluators should use a combination of techniques. This aids them to make a comprehensive search for outcomes. It also helps them cross-check the various findings. The following discussion illustrates the range of techniques that evaluators might employ.

Evaluators might assess students' test scores compared with a specified standard. The standard might be a profile of previously assessed needs, pretest scores, selected norms, program goals, or a comparison group's performance. Sanders and Horn (1994) advocate a general goal of sustained academic growth for each student, across three or more years. Webster, Mendro, and Almaguer (1994) propose comparing schools on one-year, schoolwide gains, when student background variances have been partialed out. The evaluators might use published objective tests or specially made criterion-referenced tests. They might also employ performance assessments. Experts might compare program recipients' work products against their previously assessed needs.

To assess outcomes that extend beyond an enterprise's goals, evaluators need to search for unanticipated outcomes, both positive and negative. They might conduct hearings or group interviews to generate hypotheses about the full range of outcomes and follow these up with clinical investigations intended to confirm or disconfirm the hypotheses. They might conduct case studies of the experiences of a carefully selected sample of participants to obtain an in-depth view of the program's effects.

They might survey, via telephone or mail, a sample of participants to obtain their judgments of the service and their views of both positive and negative findings. They might ask these respondents to submit concrete examples of how the project or other service influenced their work or well-being, either positively or negatively. These could be written pieces, other work products, new job status, or negative consequences. They might engage observers to identify what they believe to be program and comparison groups' achievements. They can then use the reported achievements to develop tests that reflect the hypothesized outcomes. By administering the test to program recipients and a comparison group, the evaluators can estimate the intervention's unique contributions that possibly are remote from the intended outcomes (see Brickell, 1976).

Evaluators might also conduct a "goal-free evaluation" (Scriven, 1991). Accordingly, the evaluator engages an investigator to find whatever effects an

intervention produced. The evaluator purposely does not inform the goal-free investigator about the intervention's goals. The point is to prevent the investigator from developing tunnel vision focused on stated goals. The evaluator then contrasts identified effects with the program beneficiaries' assessed needs. This provides a unique approach to assessing the intervention's merit and worth, whatever its goals.

Reporting of product evaluation findings may occur at different stages. Evaluators may submit interim reports during each program cycle. These should show the extent the intervention is addressing and meeting targeted needs. End-of-cycle reports may sum up the results achieved. Such reports should interpret the results in the light of assessed needs, costs incurred, and execution of the plan. Evaluators may also submit follow-up reports to assess long-term outcomes.

People use product evaluations to decide whether a given program, project, service, or other enterprise is worth continuing, repeating, and/or extending to other settings. A product evaluation should provide direction for modifying the enterprise or replacing it so that the institution will more cost-effectively serve the needs of all intended beneficiaries. It might also help potential adopters decide whether the approach merits their serious consideration.

Product evaluations have psychological implications, since by showing signs of growth and/or superiority to competing approaches, they reinforce the efforts of both staff and program recipients. Likewise, they may dampen enthusiasm and reduce motivation when the results are poor. The latter point brings to mind the important caveat that product evaluation reported too early in an innovative project can intimidate staff and stifle their creativity. Evaluators should be sensitive to this possibility and avoid premature feedback of possibly chilling product evaluation findings.

Product evaluation information is an essential component of an accountability report. When authorities document significant achievements, they can better convince community and funding organizations to provide additional financial and political support. When authorities learn that the intervention made no important gains they can cancel the investment. This frees funds for more worthy interventions. Moreover, other developers can use the product evaluation report to help decide whether to pursue a similar course of action.

Work Sample Technique as Applied to Product Evaluations

Del Schalock and a team at Western Oregon University (Schalock, Schalock, & Girod, 1998) are employing the *Work Sample Technique* to evaluate student teachers. They require each student teacher to develop and apply a work sample assessment exercise keyed to an instructional unit's goals. Work samples are supposed to give the student clear learning goals and performance exercises for showing mastery of the goals. A student teacher develops a work sample according to specifications and administers it to each student before instruction and following instruction. Teachers might employ a parallel form of the work

sample after instruction to help reduce effects of teaching the test. The supervisor then examines pretest-posttest gains for each part of the work sample. They do so at the level of individual students; at the level of high, medium, and low ability groups; and overall. The teacher and his or her supervisor then carefully examine the results. They assess the teacher's effectiveness in helping every student achieve the learning goals. They also assess the validity of the teacher's assessment exercises. Supervisors use these assessments to help teachers gauge teaching competence, set improvement goals, and improve their abilities to prepare classroom assessment materials.

The work sample product evaluation technique is strong in instructional validity. It directly reflects instructional goals. It helps the teacher determine whether students mastered the learning goals and how much they gained. The technique also helps the teacher develop facility in developing and using performance assessments keyed to instruction.

Bonuses of using the technique are that it provides a basis for examining whether teachers are:

- teaching and assessing high- or low-level goals
- proficient in developing high quality assessment devices that reflect the goals
- effective in teaching their students
- equally effective in teaching all levels of students

However, a cautionary note is in order. Using the Work Sample Technique to support high stakes decisions – e.g., state licensing – is controversial. The technique has not shown sufficient reliability and validity to warrant such use (Stufflebeam, 1997). Also, to use it in high stakes decision making undoubtedly would cause teachers to employ the technique to meet state expectations and thus teach the test. Nothing would prevent them from working with students to fabricate gains data. Evaluators have reported such undesirable outcomes in high stakes testing; these occurred even with more rigorous controls than the Work Sample Technique provides (House, Rivers, & Stufflebeam, 1974; Pedulla, Haney, Madaus, Stufflebeam, & Linn, 1987; Stufflebeam, Nitko, & Fenster, 1995). Under low stakes conditions, work samples are valuable classroom assessment tools. Hopefully, users will not destroy the technique's utility for teacher development and classroom assessment by placing it in a threatening, high risk context. Predictably, this would cause some teachers to cheat in showing good results and discourage others from using the technique for formative classroom assessment purposes.

Continuous Progress Matrix Sampling Testing Technique as Used in Product Evaluations

The *Continuous Progress Matrix Sampling Testing Technique* is a product evaluation technique that I use in classroom teaching. This technique provides a

periodic look at a course's evolving gross learning product and students' progress and retention of each course unit. The technique is designed to help teachers and students overcome their frequent dissatisfaction with pretest-posttest gains data. These indicate only what students gained over several months; they do not show what learning trends occurred between the two tests. Instructors express frustration when the gains are small; they do not know why, and they learned this too late to do anything about it. Probably most instructors and students would be interested to see and examine learning trends between a pretest and posttest. Then they could decide to revisit important content that the students either did not learn or did not retain.

The *Continuous Progress Matrix Sampling Testing Technique* is based on matrix sample testing (Cook & Stufflebeam, 1967; Owens & Stufflebeam, 1964). An instructor administers a parallel form of the final course examination about weekly. The different students are randomly assigned to complete different, small random samples of the test items. The instructor analyzes the results to maintain week-by-week trend lines for the total test and each course unit. During selected class sessions the instructor devotes only about the first five minutes to administering the test. This is possible since each student completes only a small sample of the total set of test questions. Starting with the second class session, the instructor distributes and explains the latest update on trends in tested achievement.

Each week, the instructor and students can see how well the class as a whole is progressing toward a high score on the final exam. By looking at the trend line for the unit taught last, the students can see whether they, as a group, mastered the material. They can also assess whether they retained or regressed in what they learned in units taught earlier. Instructors are encouraged when they see that test scores for previously untaught units remained, week after week, at the chance level, then dramatically improved following instruction. They should be concerned when test score trends show that students regressed on previously mastered material. Such feedback can motivate instructors and students to revisit and regain the prior learning. It can lead instructors to search for a better way to teach the material. Students and the instructor can discuss the results weekly to detect where past instruction and learning activities may have been weak and for what reasons. They can collaborate in deciding what material they should review and how the instructor could best get it across. This technique employs an educationally sound approach to teaching the test.

Advantages of this approach are that it helps students see that:

- testing in the course is instrumental to improving teaching and learning
- they are partners in producing a good outcome for the entire class
- they and the instructor can use relevant empirical data to assess progress and recycle instructional and learning plans
- the time involved in taking weekly tests can be small
- weekly testing is not threatening since students receive no individual scores

Limitations of the technique are that it:

- provides no feedback on performance of individual students
- is based exclusively on multiple choice test questions
- obtains feedback on each item from only one or a few students(s)

Overall, the technique is decidedly better than a pretest-posttest or posttest only approach. Like these approaches, it assesses course effectiveness. Equally or more important, it also guides instruction and learning activities. It also reviews week-to-week (or day-to-day) learning trends for each part of the course and for the overall course. Of special note, it engages the students and instructor as collaborators in using evaluation feedback constructively and continually to strengthen a course.

DESIGNING EVALUATIONS

Once the evaluator and client have decided to conduct a context, input, process, or product evaluation (or some combination), the evaluator needs to design the needed work. This involves preparing the preliminary plans and subsequently modifying and explicating them as the evaluation proceeds. Decisions about such evaluation activities form the basis for contracting and financing the evaluation, working out protocols with the involved institutions, staffing the evaluation, scheduling and guiding staff activities, and assessing the evaluation plans. The design process also provides opportunities for developing rapport, effecting communication, and involving the evaluation's stakeholder groups.

Table 3 outlines points to be considered in designing an evaluation. These points are applicable when developing the initial design or later when revising or explicating it.

The formulation of the design requires that the client and evaluators collaborate, from the outset, when they must agree on a charge. The client needs to identify the course, project, program, institution, or other object they will evaluate. The evaluator should help the client define clear and realistic boundaries for the study. The client is a prime source for identifying the various groups to be served by the evaluation and projecting how they would use it. The evaluator should ask clarifying questions to sort out different (perhaps conflicting) purposes. They should also get the client to assign priorities to different evaluation questions. The evaluator should recommend the most appropriate general type(s) of study (context, input, process, and/or product). The client should confirm this general choice or help to modify it. In rounding out the charge, the evaluator should emphasize that the evaluation should meet professional standards for sound evaluations.

The evaluator should define the data collection plan. He or she should provide an overview of the general evaluation strategies. These could include surveys, case studies, site visits, advocacy teams, goal-free searches for effects, adversary hearings, a field experiment, etc. The evaluator should also write technical plans for collecting, organizing, and analyzing the needed information. He or she

Table 3. Outline for Documenting Evaluation Designs

Review of the Charge
- Identification of the course or other object of the evaluation
- Identification of the client, intended users, and other right-to-know audiences
- Purpose(s) of the evaluation (i.e., program improvement, accountability, dissemination, and/or increased understanding of the involved phenomena)
- Type of evaluation (e.g., context, input, process, or product)
- Values and criteria (i.e., basic societal values, merit and worth, CIPP criteria, institutional values, technical standards, duties of personnel, and ground-level criteria)
- Principles of sound evaluation (e.g., utility, feasibility, propriety, and accuracy) to be observed

Plan for Obtaining Information
- The general strategy (e.g., survey, case study, advocacy teams, or field experiment)
- Working assumptions to guide measurement, analysis, and interpretation
- Collection of information (i.e., sampling, instrumentation, data collection procedures and instruments, and permissions from data sources)
- Organization of information (i.e., coding, filing, and retrieving)
- Analysis of information (both qualitative and quantitative)
- Interpretation of findings (i.e., interpretive standards, processing judgments, developing conclusions)

Plan for Reporting the Results
- Drafting of reports
- Prerelease reviews and finalization of reports
- Dissemination of reports
- Provision for follow-up activities to assist uses of the evaluation
- Plan for responding to anticipated attacks on the evaluation

Plan for Administering the Evaluation
- Summary of the evaluation schedule
- Plan for meeting staff and resource requirements
- Provision for metaevaluation
- Provision for periodic updates of the evaluation design
- Budget
- Memorandum of agreement or contract

should obtain and consider stakeholders' reactions to the data collection plan. The evaluator and client should anticipate that the data collection plan will likely change and expand during the evaluation. This will happen as they identify new audiences and as information requirements evolve.

Evaluators should gear reporting plans to achieve use of the evaluation findings. They should involve clients and other audiences in deciding the contents and timing of needed reports. Stakeholders should also help in planning how the evaluator will disseminate the findings. The reporting plan should consider report formats and contents, audiovisual supports, review and revision, means of presentation, and right-to-know audiences. Appropriate procedures to promote use of findings might include oral reports and hearings, multiple reports targeted to specified audiences, press releases, sociodramas to portray and explore the findings, and feedback workshops aimed at applying the findings. The client and evaluator should seriously consider whether the evaluator might play an important role beyond the delivery of the final report. For example, the client might engage the evaluator to conduct follow-up workshops on applying the findings.

Such followup work can be as important for helping audiences avoid misinterpretation and misuse of findings as for helping them understand and make appropriate use of the results. Also, only neophyte evaluators are surprised when some person(s) or group(s) that don't like the evaluation's message attack and otherwise try to discredit the work. Throughout the design and reporting processes evaluators should be sensitive to the politics attending the evaluation and make tentative plans to address unwarranted and unfair attacks on the evaluation.

The final part of the design is the plan for administering the evaluation. The evaluator should identify and schedule the evaluation tasks consistent with the needs of the client and other audiences for reports and in consideration of the relevant practical constraints. The evaluator needs to define staff assignments and needed special resources. The latter might include office space and computer hardware and software. He or she should also assure that the proposed evaluation personnel will be credible to the program's stakeholders. The evaluator and client need to agree on who will assess the evaluation plans, processes, and reports against appropriate standards. They also should agree on a mechanism by which to periodically review, update, and document the evolving evaluation design. They need to lay out a realistic budget. Also, they should formalize contractual agreements including authority for editing and releasing findings and rules for terminating the agreement.

The discussion of Table 3 has been necessarily general, but it shows that designing an evaluation is a complex and ongoing task. It recommends that the evaluator should continually communicate with the client and other audiences and emphasizes the importance of evolving the evaluation design to serve emerging information requirements. Also, it stresses the need to maintain professional integrity and contractual viability in the evaluation work. Readers are referred to www.wmich.edu/evalctr/checklists/, where they can find a collection of checklists to use in designing and contracting various kinds of evaluations.

CONCLUSION

This chapter has presented the CIPP Evaluation Model, which provides direction for evaluations of context, inputs, process, and products. The chapter describes the CIPP Model's role in improving, researching, disseminating, and accounting for school programs and other evaluands; explains its main concepts; discusses its uses for guiding improvement efforts and for accountability; provides illustrations of application; describes techniques particularly suited to the model; and outlines the elements of sound evaluation designs. The CIPP Model is shown to be adaptable and widely applicable in many areas, including elementary, secondary, and higher education. It is recommended for use by individual educators, groups of educators, schools, and systems of schools and similar groups in disciplines outside education. Evaluators are advised to validly assess the merit of a program, service, product, or institution and determine its worth in serving all the rightful beneficiaries. The chapter's key themes are that

(1) evaluation involves assessing something's merit and worth; (2) the most important purpose of evaluation is not to prove, but to improve; (3) evaluations should be both proactive in guiding improvements and retroactive in producing accountability reports; (4) evaluators should assess goals, strategies, plans, activities, and outcomes; (5) evaluations should be grounded in sound, clear values; (6) evaluators should be interactive in effectively communicating with and serving clients and other right-to-know audiences; (7) evaluation design and reporting are ongoing processes that should be tailored to meeting the audiences' information needs; (8) evaluators should be sensitive to and appropriately resistant to attempts by persons or groups to corrupt or discredit the evaluation; (9) a program's success should be judged on how well it meets the assessed needs of targeted beneficiaries; (10) evaluations should employ multiple approaches to gather relevant information, including both quantitative and qualitative methods; (11) whatever the methods employed, the evaluation should meet appropriate standards for sound evaluations; and (12) evaluations themselves should be evaluated through internal and external metaevaluations.

APPENDIX:

EXAMPLES OF EVALUATIONS THAT WERE GUIDED BY THE CIPP EVALUATION MODEL

The following is a sampling of evaluations conducted by the Western Michigan University Evaluation Center. In varying degrees, these evaluations illustrate The Center's wide ranging use of the CIPP Evaluation Model. They are listed here at the section editor's recommendation. He suggested that this chapter's readers would be interested to know what kinds of applications of the CIPP Evaluation Model they could expect to learn about at the Evaluation Center. At a minimum, the following list conveys the variety of uses to which the CIPP Model has been put. While the model was initiated in education, the following examples show that it can be applied to a wide range of settings and content.

Community Development

1. evaluation of Consuelo Zobel Alger Foundation's self-help housing program in Hawaii
2. external evaluation of the MacArthur Foundation-sponsored Fund for Community Development that assisted selected community development corporations in Chicago to improve housing, commerce, and industry in their neighborhoods
3. evaluation of the Standard Oil weatherization program in Cleveland

Higher Education

4. an evaluation of the Hill Family Foundation's program to improve productivity in higher education
5. a technical assistance project to aid Western Michigan University to develop a universitywide system of program review
6. evaluation of the Mott Foundation's Program for the Historically Black Colleges
7. evaluation of the Western Michigan University College of Education's external doctoral program in Guam

International

8. evaluation of Consuelo Foundation's socialized housing project in Negros, Philippines
9. evaluation for the World Bank of teacher education in the Philippines

Personnel Evaluation

10. evaluation and design of a personnel evaluation system for the U.S. Marine Corps
11. evaluation and advice for improving teacher evaluation systems in Hawaii, Alaska, and Ohio
12. development of an evaluation criteria shell to guide the National Board for Professional Teaching Standards' development and validation of assessment systems to certify highly accomplished K-12 teachers

Schools and the Personnel

13. evaluation of charter school initiatives (Michigan, California, Connecticut, Ohio, Pennsylvania, and Illinois)
14. evaluation of various aspects of K-12 schools including curricula, extended year programs, and community perceptions of schools
15. evaluation of Goals 2000 and Technology Literacy Challenge Fund projects in Michigan
16. a study for the National Science Foundation of the effects of the 1977 energy crisis on Columbus, Ohio, public schools
17. evaluation of the Alabama educator Inservice Centers
18. evaluation of the Indianapolis School Partners program for Lilly Endowment, Inc.
19. program evaluations for the Michigan Partnership for New Education

Science Education

20. evaluations for the National Science Foundation of Delaware and Oregon system projects in science education and mathematics education
21. evaluation of the impact and effectiveness of the National Science Foundation's Advanced Technological Education (ATE) program
22. evaluation of the National Science Foundation-sponsored Rural Systemic Initiatives Program
23. evaluation of educational programs for the Environmental Protection Agency
24. evaluation of science education training provided by the Argonne National Laboratory

Social/Youth

25. evaluation of Michigan's Life Services project for coordinating welfare services
26. evaluation of Michigan programs in supported employment, housing, and transition from school to work
27. evaluation of the W.K. Kellogg Foundation-sponsored Kellogg Youth Initiatives Program
28. evaluation of gambling addiction in Michigan for the Michigan Lottery
29. survey of female athletes in Michigan high schools about the possible realignment of high school sports seasons to conform to intercollegiate seasons

State/Regional Educational Services

30. evaluation of Appalachia Educational Laboratory programs
31. development of an evaluation system for Ohio's state system for career education
32. evaluation of Michigan's regional educational media centers
33. evaluation of the research and development departments of government and educational organizations

Testing

34. evaluation of the Michigan Educational Assessment Program
35. evaluation of the Kentucky Instructional Results Information System

Metaevaluation

36. metaevaluation of seven undergraduate engineering programs
37. metaevaluation of Australia's national distance baccalaureate program

38. metaevaluation of the National Assessment Governing Board's attempt to set achievement levels on the National Assessment of Educational Progress
39. metaevaluation of the teacher education program at St. Patrick's College, Dublin, Ireland
40. metaevaluation of teacher evaluation and school accountability for Texas
41. metaevaluation of an evaluation of the New York City school district's testing of the Waterford Integrated Learning System-a computer-based skills program for elementary school students

ENDNOTES

[1] CIPP Model has withstood the test of time and practice over many years. The chapter's appendix lists examples of the wide variety of evaluations that employed this model.

[2] The Evaluation Center was established in 1963 at The Ohio State University and has been at Western Michigan University since 1973. Since its inception The Evaluation Center has conducted a wide range of projects aimed at advancing the theory and practice of evaluation and has provided a learning laboratory for many graduate students, visiting scholars, and practitioners. It is the home base of the North American Joint Committee on Standards for Educational Evaluation and from 1990 through 1995 housed the federally funded national research and development center on teacher evaluation and educational accountability. Among the Center's experiences are applications in elementary and secondary education, continuing medical education, community and economic development, self-help housing, community programs for children and youth, administrator evaluation, and military personnel evaluation.

REFERENCES

Alexander, D. (1974). *Handbook for traveling observers*, National Science Foundation systems project. Kalamazoo: Western Michigan University Evaluation Center.

Brickell, H.M. (1976). Needed: Instruments as good as our eyes. *Occasional Paper Series, #7.* Kalamazoo: Western Michigan University Evaluation Center.

Cook, D.L., & Stufflebeam, D.L. (1967). Estimating test norms from variable size item and examinee samples. *Educational and Psychological Measurement*, **27**, 601–610.

Evers, J. (1980). *A field study of goal-based and goal-free evaluation techniques*. Unpublished doctoral dissertation. Kalamazoo: Western Michigan University.

House, E.R, Rivers, W., & Stufflebeam, D.L. (1974). An assessment of the Michigan accountability system. *Phi Delta Kappan*, **60**(10).

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.

Manning, P.R., & DeBakey, L. (1987). *Medicine: Preserving the passion*. New York: Springer-Verlag.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative of Educational Reform*. Washington, DC: U.S. Government Printing Office. (No. 065-000-00177-2).

Nowakowski, A. (1974). *Handbook for traveling observers*, National Science Foundation systems project. Kalamazoo: Western Michigan University Evaluation Center.

Pedulla, J.J., Haney, W., Madaus, G.F., Stufflebeam D.L., & Linn. R L. (1987). Response to assessing the assessment of the KEST. *Kentucky School Boards Association Journal*, **6**(1), 7–9.

Owens, T.R., & Stufflebeam, D.L. (1964). An experimental comparison of item sampling and examinee sampling for estimating test norms. *Journal of Educational Measurement*, **6**(2), 75–82.

Reed, M. (1989). *Fifth edition WMU traveling observer handbook: MacArthur project*. Kalamazoo: Western Michigan University Evaluation Center.

Reed, M. (1991). *The evolution of the traveling observer (TO) role*. Presented at the annual meeting of the American Educational Research Association, Chicago.

Reinhard, D. (1972). *Methodology development for input evaluation using advocate and design teams.* Unpublished doctoral dissertation, The Ohio State University, Columbus.

Sandberg, J. (1986). *Alabama educator inservice training centers traveling observer handbook*. Kalamazoo: Western Michigan University Evaluation Center.

Sanders, W., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, **8**(3), 299–312.

Schalock, D., Schalock, M., & Girod, J. (1997). Teacher work sample methodology as used at Western Oregon State College. In J. Millman (Ed.), *Using student achievement to evaluate teachers and schools* (pp. 15–45). Newbury Park, CA: Corwin.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Stufflebeam, D.L. (1997). The Oregon work sample methodology: Educational policy review. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 53–61). Thousand Oaks, CA: Corwin.

Stufflebeam, D.L., Nitko, A., & Fenster, M. (1995). *An independent evaluation of the Kentucky Instructional Results Information System (KIRIS).* Kalamazoo: Western Michigan University Evaluation Center.

Sumida, J. (1994). *The Waianae self-help housing initiative: Ke Aka Ho'ona: Traveling observer handbook.* Kalamazoo: Western Michigan University Evaluation Center.

Thompson, T.L. (1986). *Final synthesis report of the life services project traveling observer procedure.* Kalamazoo: Western Michigan University Evaluation Center.

Webster, W.J., Mendro, R.L., & Almaguer, T.O. (1994). Effectiveness indices: A "value added" approach to measuring school effect. *Studies in Educational Evaluation*, **20**(1), 113–146.