

STRATEGIES FOR INSTITUTIONALIZING EVALUATION: REVISITED

Abstract

Every service organization needs to conduct sound evaluations to identify and address constituents' needs, improve services, make defensible personnel decisions, effectively serve clients, and earn client confidence. This article posits that an organization can best meet its evaluation needs by institutionalizing a sound unified evaluation system. While the article focuses on educational organizations, its message applies to the full range of organizations dedicated to serving clients. To assist organizations to define one general approach to program, client, and personnel evaluation, the article presents 2 checklists. The first defines 18 goals of a sound, unified evaluation system. The second checklist defines 10 components of a fully functional evaluation system. Organizations are advised to use these checklists to examine and strengthen or replace their existing evaluation systems.

by

Daniel L. Stufflebeam
The Evaluation Center
Western Michigan University

May 1997

Editor's Note:

This volume of the Occasional Papers Series (OPS) might alternately have been titled the lost manuscript. It was completed shortly before the author, Daniel Stufflebeam, was involved in a serious auto pedestrian accident. After nearly three years, the manuscript was rediscovered and submitted for publication in the OPS series.

The main thesis of this article is that organizations should institutionalize a unified evaluation system in order to help identify client needs, improve services, make defensible personnel decisions, etc. This is a lesson that is important for all organizations to learn. It is with pride that we present this fine paper as the 18th volume in OPS series.

Craig Russon
Editor
The Occasional Papers Series

STRATEGIES FOR INSTITUTIONALIZING EVALUATION: REVISITED^{1, 2}

Background

In 1970, Egon Guba and I collaborated on a paper titled *Strategies for the Institutionalization of the CIPP Evaluation Model*. We proposed that school districts and other educational organizations install a unified evaluation approach that would efficiently address the many accountability requirements then being pressed on them and also regularly supply information for planning and guiding projects. The paper emphasized that external evaluation services, while important, could not fully serve educational organizations, because there were far too few evaluation centers, companies, and consultants to address the evaluation needs of about 16,000 school districts, 50 state education departments, and hundreds of other educational organizations. Organizations needed an ongoing process of internal evaluation to help the staff to constantly learn from experiences and improve practices and regularly to report accomplishments to sponsors and other external audiences. Moreover, internal evaluations would serve an organization better if all members subscribed to the same sound view of evaluation. We offered the CIPP³ model as a comprehensive framework of pertinent questions and information that diverse groups could use as a common evaluation philosophy and language. We then examined the model's requirements to determine what steps an organization needs to follow in developing and employing a fully functional evaluation system. In general, we advised school districts and other educational organizations to empower

¹ This article was supposed to help address the “postmodern dilemma” in evaluation and help evaluation theorists address their identity crises. In my experience, the most important current dilemma in evaluation is that many service organizations do an inexcusably poor job of conducting and using evaluations to improve services. Thus, this article is targeted not to help evaluation authors address their worries over such matters as postmodernism, poststructuralism, empowerment, constructivism, and the nature of truth, but to provide conceptual and practical guidance to organizations that want and need to install a sound system of improvement-oriented evaluation.

² The article reflects lessons learned from a wide range of evaluation and metaevaluation experiences of the Western Michigan University Evaluation Center. These focused on such enterprises as Open Learning Australia; the U.S. National Assessment of Educational Progress; the Kentucky Instructional Results Information System; the New York City school district's Integrated Learning System; the Hawaii Comprehensive Assessment and Accountability System; the Oregon, Ohio, and Texas systems for evaluating teachers; Teach for America; the Dallas, Des Moines, and many other school district evaluation departments; the U.S. Marine Corps personnel evaluation system; two National Science Foundation state systems projects; and community development and youth programs in Michigan, Hawaii, Poland, Mexico, and the Philippines.

³ The **CIPP Model** posits that evaluations should assess: **CONTEXT** to focus efforts on assessed client needs, improvement opportunities, and problems to be solved; **INPUTS** to examine and help improve plans for addressing priority needs; **PROCESS** to record, judge, and guide implementation of plans; and **PRODUCT** to identify and judge intended and unintended outcomes. CIPP evaluations are used proactively to guide decisions and retroactively to issue summatively-oriented accountability reports. The acronym CIPP derives from the first letters of the CIPP Model's four types of evaluation.

themselves by institutionalizing a systematic process of evaluation that would inform the organization's ongoing decision process, help assure its accountability to funders and constituents, and complement and cross-check external evaluations of the organization's contributions.

Dr Guba and/or I were privileged to help a number of different types of educational organizations successfully carry out our evaluation advice. These included school districts (e.g., Columbus, Xenia, and Cincinnati, Ohio; Saginaw, Lansing, and Detroit, Michigan; Dallas, Fort Worth, and Houston, Texas; and Des Moines, Iowa), regional educational laboratories (e.g., the Austin Southwest Educational Development Laboratory and the Northwest Regional Educational Laboratory); state education departments (including Montana, Michigan, and Ohio), research and development centers (e.g., the National Center for Vocational and Technical Education and the Wisconsin R and D Center on Student Learning), and the U.S. Office of Education.

The 1970 article could draw on only modest developments within the evaluation field compared to the vast evaluation experience and progress considered in this paper. The more recent relevant developments include

- professional societies of evaluators with their attendant professional journals
- professional standards for both program evaluation and personnel evaluation
- a large and growing literature of evaluation
- evaluation degree programs and a small, but steady supply of well-trained graduates
- organizations that have successfully institutionalized an evaluation system
- organizations that have successfully integrated program, client, and personnel evaluations
- extensive experience in conducting metaevaluations
- evaluation models that have been tested and validated in practice
- strong progress toward consensus that quantitative and qualitative approaches are complementary, not contradictory
- more realistic attitudes about the limitations as well as the potential contributions of such inquiry procedures as experimental design and standardized testing

Clearly, today's organizations can draw on a much richer supply of evaluation experience, ideas, personnel, and tools than was the case in the early 1970s.

This article affirms what Guba and I recommended 27 years ago: that **every service organization should adopt and apply a shared and valid concept of evaluation.** In addition, the article stresses that the common approach to evaluation should be applicable to all the organization's evaluations. These include evaluations of client needs and performance, personnel competence and performance, and program effectiveness, also evaluations at the different levels of the organization.

It seems logical that a unified and fully functional evaluation system could be one of an organization's most effective tools. It would provide the organization's personnel with a common process and language for study, problem solving, negotiation, decision making, and reporting to the public and higher levels of authority. It would provide information on the full range of issues in the organization. Among others, these include client needs, personnel

qualifications and performance, service plans and budgets, program/service implementation and costs, organizational accomplishments, day-to-day progress of individual students or other clients, and community perceptions and support of the organization. A sound organizational evaluation system would provide feedback throughout the year and not just at the end. It would be applied at all levels: policy and administration, departments, work groups, and individual staff members. It would give the organization and each staff member one general model to evaluate personnel, programs, and services. It would stress that evaluations should be designed and used for both improvement and accountability.

The remainder of this article describes a general approach that an organization could use to evaluate all its important functions. This article's recommended general approach to evaluation addresses two questions:

1. What is an appropriate vision for organizational evaluation?
2. What steps can an organization take to realize the vision?

These questions are addressed by two checklists designed to help organizations conceptualize evaluation and plan evaluation systems. The Type 1 Checklist lists 18 requirements of a unified concept of organizational evaluation. The Type 2 Checklist identifies 10 components of a fully functional organizational evaluation system. In short, the Type 1 Checklist defines goals for a sound evaluation approach and the Type 2 Checklist defines the components or mechanisms required to achieve the goals.

These are generic checklists. Organizations should use them as general guides, not prescriptions. The point needs to be underscored that **an organization's administration should not "lay on" a common approach to evaluation**. They should work it out with representatives of all the stakeholders. The two checklists are offered as tools to help an organization's stakeholders guide their deliberations in the course of defining and designing a unified approach to evaluation.

To make use of this article, organizations are advised to begin by appointing an evaluation planning team. It should be representative of the organization's personnel and constituents. The organization should assign the team to assess the organization's present evaluation system and devise a plan for improving or replacing it.

Type 1 Checklist: Requirements for a Unified Concept of Evaluation

Exhibit 1 contains the Type 1 Checklist. In using this checklist, the evaluation planning team should examine each checkpoint to determine whether it conforms to the organization's philosophy. Then they should revise, drop, or replace the checkpoints as appropriate. They should finalize the checklist so that it is consistent with the organization's values and responsive to needs to improve the current evaluation system. They should also make sure they can defend the soundness of the revised checklist. Through this process, the evaluation planning team can establish the conceptual foundation needed to improve or replace the organization's evaluation system.

EXHIBIT 1
TYPE 1 CHECKLIST:
REQUIREMENTS OF A UNIFIED
CONCEPT OF EVALUATION

1. PROMOTE SERVICE TO ALL TARGETED CLIENTS
2. ASSESS ALL COMPONENTS THAT INFLUENCE ORGANIZATIONAL SUCCESS
3. MAINTAIN CONTINUITY & FLEXIBILITY
4. WORKABLE IN THE ORGANIZATION
5. UNDERSTOOD & VALUED BY ALL PERSONNEL
6. ASSESS PERFORMANCE OF INDIVIDUALS, GROUPS, & THE OVERALL ORGANIZATION
7. FOSTER IMPROVEMENT
8. SUPPLY ACCOUNTABILITY REPORTS
9. ENCOMPASS THE RELEVANT CRITERIA
10. PROMOTE EXCELLENCE
11. PROMOTE EQUALITY OF OPPORTUNITY
12. BALANCE IN ASSESSING PROGRAMS AND PERSONNEL
13. RECIPROCITY IN ASSESSING ALL STAFF
14. PARALLELISM IN RIGOR & FREQUENCY OF ASSESSMENT
15. EFFICIENCY
16. CREDIBLE TO STAKEHOLDERS
17. LEGALLY VIABLE & FAIR
18. MEET PROFESSIONAL REQUIREMENTS FOR SOUND EVALUATION

Below are explanations of each of the 18 checkpoints.

1. The evaluation system should promote service to the full range of targeted clients.

The organization should periodically clarify its intended customers and examine their needs for organizational services. Its evaluation system should help the organization's staff to examine customer gains and satisfaction. For example, a school district should examine achievements of all its students, including different ethnic and ability groups and students with handicapping conditions. The evaluation feedback should help the organization assure that it is effectively addressing the targeted needs of all the customers. It should help the individual service providers, such as classroom teachers, monitor needs, problems, and progress of all their students or other targeted clients. No public service organization, such as a school district, should delete this checkpoint. It might clarify, expand, or otherwise strengthen the checkpoint. But it must not unintentionally or intentionally underserve any subgroup of constituents.

2. Assess all components of the organization that influence organizational success.

Some organizations evaluate their programs and services based mainly or only on outcomes, such as student test scores. This approach is much too narrow, since it omits organizational processes employed to produce outcomes. An organization's evaluation system must look at every component that influences organizational success. Among others, these are staff qualifications and performance, program plans and the planning process, materials and equipment, uses of technology, staff participation in decision making, staff development, the organizational calendar and work schedule, publicity and communications, leadership and supervision, internal evaluation, and organizational policies.

Also, an approach that only assesses achievement of organizational goals is too narrow in its examination of outcomes. This is so because it fails to look for important positive and negative side effects and also may fail to assess customer benefits regarding the full range of their needs for service.

The consequences of a too narrow approach to evaluation are especially clear and important in the case of school districts. For example, districts that assess only reading and math scores miss important achievements concerned with physical, emotional, aesthetic, moral, vocational, social, and intellectual development. Also, testing programs usually report student scores too late to guide instruction during the involved school year. Moreover, an approach based exclusively on student test scores can lead to erroneous judgments. If the scores are low, the district might blame the students' family backgrounds, even if the real cause is poor instruction. If the scores are high, the district might take credit, even though the main cause might be that its students come from families that are overcoming the district's poor service. For example, Americans often give credit to schools in Taiwan and other countries whose students outscore U. S. students on mathematics and other subject area tests, without considering that the majority of the students in those countries may spend about as much time with special tutors, preparing for national tests, as they do in the schools. Clearly, feedback that is limited to student outcomes does not provide an adequate basis for understanding, judging, and improving the effectiveness of a school district's services.

3. Maintain both continuity and flexibility in the evaluation system.

Organizations must maintain continuity in collecting data for longitudinal studies in order to be accountable for improvements from year to year. Fulfillment of this requirement is exemplified by the Tennessee state education department's system for tracking and analyzing year-to-year gains for individual students, as well as students grouped by classroom, school, and district (Sanders & Horn, 1993). Similar value-added assessment systems are being operated successfully in the Dallas Independent School District (Webster, Mendro, & Almaguer, 1993, 1994) and in Great Britain (Tymms, 1995).

In addition to meeting the continuity requirement, organizations must also obtain information that responds to emergent questions. An organization should therefore allow some flexibility in its evaluation system. Annually, it should reassess evaluation priorities and allocate evaluation resources accordingly. The Dallas Independent School District exemplifies this point through its engagement of a standing Accountability Commission charged to annually review and adjust the District's evaluation priorities and define evaluation indicators (Webster, 1994).

4. Employ an evaluation concept that will work in the particular organization.

An organization should allocate sufficient staff time and budget so that its evaluation system will strongly support decision making and accountability at all levels. It should provide the organization's leaders and staff members with direction for strengthening programs/services and information for informing the organization's policymakers, funders, clients, and others about the quality and effectiveness of services.

However, the organization must keep its evaluation activities within bounds of feasibility. Often an organization must choose low cost methods to obtain the needed information. Whenever possible, it should incorporate external evaluations of the organization or organizational components into its own evaluation system. Using data from external evaluations, such as accreditation studies, can ease the burden for locally generated data. Finally, the organization should seek ways to build evaluation into the roles of individual staff members and committees, so that they integrate evaluation into their day-to-day activities. As much as possible an organization should make evaluation a regular part of its policy development, planning, service delivery, supervision, staff development, and reporting to constituents.

5. Adopt and define a concept of evaluation that the organization's leaders and staff understand and value.

The organization's personnel must understand, accept, and apply the adopted evaluation approach if it is to function effectively. The approach must be theoretically sound, well defined, validly instrumented, and responsive to the organization's information needs. However, these attributes will matter little if the organization's personnel are neither committed nor able to make it work. The organization should involve policymakers, administrators, line and support staff members, clients, and other stakeholders in defining the organization's evaluation model. The organization should use input from these groups to adopt appropriate evaluation policies and periodically to review and improve the evaluation system.

6. Promote evaluations to examine and guide the work of individual staff members, groups, and the organization as a whole.

Most organizations have a unique hierarchy of staff and work activities. The involved levels may include the individual staff members (such as teachers), task groups (such as curriculum committees), departments and divisions, and the organization as a whole. As much as possible, the organization should foster and support evaluations at all levels. To accomplish this, the organization's leaders should help every staff member adopt an evaluation orientation. The organization should train its personnel to collect, analyze, and use data to guide decisions and prepare accountability reports. It should help them to think critically about the work and how they might constantly improve it. As feasible, the organization should also provide staff members with technical support and useful evaluation materials.

7. Employ evaluation for improvement.

Evaluation is not important in its own right. Its most important function is to help organizations improve services to clients. No evaluation system can be of much use in the improvement process if it only provides feedback after the fact. Organizations must conduct evaluations proactively to guide decision making. They must integrate evaluation into all important aspects of the organization and provide personnel with ongoing feedback for improvement. Also the organization's leaders should keep in mind that an important way to improve organizational services is to identify and discard hopelessly flawed or useless activities and to dismiss persistently incompetent staff members. While this is not a pleasant line of thought, it is an essential one if organizations are serious about providing their clients with excellent service.

8. Supply accountability reports.

Despite the importance of an improvement orientation, an organization's evaluation system must also provide its constituents with accountability reports. Funding agencies, regulatory groups, accrediting organizations, taxpayers and other constituents desire and have a right to receive information on the organization's expenditures and the quality of services and client outcomes.

An organization might meet its accountability requirements in several ways. One of the best ways is to provide evidence that the organization regularly evaluates all aspects of its operations and uses the information to guide improvement. Also, some organizations can meet part of their accountability requirements by participating in evaluations that compare their services and outcomes with similar organizations.

However, to be fair, such comparative evaluations must take account of the different organizations' special circumstances. For example, some hospitals have much higher mortality rates than others due to serving unusually large numbers of terminally ill and high risk patients. Also, when comparing schools or school districts on the test scores of their students, it is important to consider student backgrounds and community characteristics. Sometimes poor or strong student outcomes have more to do with factors outside the control of a school or district, such as educational support in the home and violence in the neighborhood, than with the quality of the school's or district's services. In such cases, comparative evaluations might be more

reflective of the students' family backgrounds and neighborhood conditions rather than the schools' or districts' levels of effectiveness in serving students.

Finally, organizations should guard against issuing accountability reports that are untruthful public relations devices. To be accountable, an organization must issue factual reports that show areas for development as well as strengths. In the long run, candid, honest reporting of problems as well as improvement efforts enhances an organization's credibility and encourages constituents to financially support or otherwise assist the organization's improvement process.

9. Invoke the full range of relevant evaluative criteria.

The root term in evaluation is value. Essentially, evaluation involves assessing services against a pertinent set of community, professional, and/or organizational values. Every organization is grounded in a complex set of values, and to some extent these vary across organizations. An organization's values provide the basis for deriving its evaluative criteria. In defining its approach to evaluation, the organization should therefore clarify its values. Moreover, in designing particular evaluations, the organization's personnel should carefully identify the pertinent criteria. The organization's values proscribe the evaluative criteria that may be selected. The chosen criteria, along with the questions of stakeholders, dictate information needs. The derived information requirements then determine what kind of evaluation instruments the organization should use.

Unfortunately, many organizations and individual evaluators approach this process backwards. They choose an available instrument that is easy to use. In such cases the instrument determines the evaluation's values, criteria, and information. Such an approach to evaluation can be worse than useless. It consumes valuable time and resources and can misdirect the evaluation. The bottom-line message here is that organizations should use values clarification as the foundation stone of their evaluation processes.

10. Promote excellence in all that the organization does.

Among the essential values in any service organization is the pursuit of excellence. This is especially true in universities, colleges, and schools. Excellence is what education is about. It is a process of helping every student develop to her or his full potential. To do this educational organizations must set and maintain high standards in everything they do. Their evaluation systems should examine whether every aspect of the organization is attaining or at least seeking to achieve and maintain a high standard of excellence. Organizational evaluation systems should promote, give credit for, and even help to celebrate excellent service and achievements.

11. Promote equality of opportunity.

Public service organizations in democratic societies share the value of equality of opportunity. For example, U. S. laws require public schools to give all their students equal access to learning opportunities and resources, irrespective of gender, ethnicity, or socioeconomic status. Similarly, public service organizations are expected to choose staff members for their abilities to carry out job responsibilities, irrespective of personal characteristics. In implementing a value of equality of opportunity, an organization's evaluation system should help the staff to search out

and address inequities. Such inequities may be seen in the organization's allocation of resources to serve different groups of clients, segregated activities, and discriminatory services and hiring practices. Sound organizational evaluation plays a vital role in helping the organization identify and address inequities and deliver equitable service.

12. Give balanced attention to assessments of clients, programs/services, and staff.

Organizations must evaluate customer gains and satisfaction to identify areas of deficiency and strength. To improve client benefits, it is often essential to improve programs and services and upgrade staff competence and performance. Thus, the organization should allocate balanced effort and resources to evaluating staff, programs, and results for clients.

13. Assure reciprocity in evaluating all staff.

Too often, organizations focus almost all their staff evaluation efforts on line personnel, such as teachers. The organization should evaluate all its personnel. It should give equitable attention to evaluating the policymakers and administrators as well as the line and support staff. All personnel have a common interest and commitment in assuring that the organization succeeds in serving its clients well. All of them can improve, no matter how well they are performing. The organization should evaluate the performance of the personnel as a whole as well as the individual members and should use the results to help personnel upgrade knowledge and skills and improve performance.

Partly, this recommendation is intended to be fair to the different groups. But the point is not to distribute blame or credit. The more important point is to evaluate and strengthen every aspect of the organization that determines its effectiveness, including the personnel. Evaluation can and should play an important role in improving the performance of every category of organizational personnel. Also, the organization must conduct personnel evaluations to reinforce and strengthen collaboration among its personnel.

This recommendation might make some board members and administrators nervous because of the prospect of embarrassing feedback. However, when they view and use evaluation as a tool for improvement, everyone gains. The evaluatees are reinforced for areas of excellent performance and gain insights about areas of deficiency and how to improve their service. Also, by participating in a continuous process of self-examination and improvement, board members and administrators set a positive, improvement-oriented example for the organization's staff and constituents.

14. Engage in parallel efforts to increase the rigor as well as the frequency and scope of evaluations.

In developing a fully functional evaluation system, an organization will likely increase the frequency and scope of its evaluations. It must not do so at the expense of rigor. Neither should an organization conduct only infrequent or narrow evaluations because of the difficulty in achieving rigor. Instead, the organization should simultaneously work to increase the frequency and scope of evaluations and to build in rigor. This means that an organization's personnel will have to study the available evaluation technology, consider options, try out some of the desired

new approaches, judiciously install new procedures that evidence rigor and, when appropriate, cut back on some of the evaluation procedures that are found to be invalid.

The state education department in Kentucky has been struggling with this issue over the past five years (Stufflebeam, Nitko, & Fenster, 1995). It attempted to launch a statewide system for evaluating schools based almost totally on performance assessments and free of multiple choice tests, also linked to rewards and sanctions to be given to individual schools. While the state expended more than 50 million dollars in developing, administering, and attempting to validate this system, it ran into great difficulties. It used various performance evaluation instrumentalities before they were validated, but did carry on a validation process. When the validation studies and external evaluations of the system showed that certain assessment components were invalid, the state had to cut out some components; launch more rigorous assessment development procedures; delay giving sanctions to schools; and return to the use of standardized, multiple choice tests. Despite its difficulties and failures, the Kentucky experience illustrates that organizations should simultaneously work at expanding the scope of measures and assuring their validity.

15. Constantly improve efficiency of evaluations.

Beyond meeting its needs for an effective evaluation system (as advocated in #14), an organization should also apply a prudential criterion. Virtually all organizations have only limited resources to invest in evaluation. They should carefully allocate these resources to obtain the most important information. They should collect this information only when they will use it. They should collect it as efficiently as possible. Evaluation is very much a human enterprise, and it is labor intensive. Organizations never can completely automate their evaluation systems. However, they can and should take all possible steps to increase the efficiency and cost-effectiveness of their evaluations.

The Kentucky experience referenced above is a case in point. As part of a statewide system of educational reform, the state decided to eliminate the use of standardized, multiple choice tests and replace them with authentic performance assessments, including portfolios, group problem solving assessments, and short answer tests. Evaluations of the performance assessments found that many of them were not only unreliable and thus invalid, but also enormously expensive in terms of money, time, and public dissatisfaction. The state is now again finding a place for standardized, multiple choice tests that evidence broader scope, higher reliability, better bases for comparisons with student achievements in other states, and lower costs than performance assessments. Valid performance assessments can make important contributions--as seen in more and improved writing by Kentucky students who have spent much time providing and receiving feedback on writing samples--but they are expensive and must be used judiciously along with other more efficient measures.

16. Earn and maintain credibility for the evaluation system.

An organization's personnel must engage in an extensive amount of self-evaluation to guide and improve their services and inform constituents about processes and outcomes. Self-evaluation is essential to professionalism, but the possibilities of self-interest, bias, and "blindness" in self-evaluations also raise questions. Thus, the organization's personnel must take

concrete steps to assure that their evaluations are beyond reproach. They must convince outsiders that this is so.

The organization can do a number of things to establish and maintain credibility for its evaluations. The organization can adopt and publicize a sound organization-wide approach to evaluation, including standards for the evaluation work. The organization can regularly train staff in evaluation principles and procedures. The organization's leaders can demonstrate in public meetings that they regularly use evaluation to guide decisions. The staff can include evaluative findings in their newsletters and other communications to constituents. Service providers should use evaluative data to help clients gauge their benefits from the organization's services. The organization can obtain information from different sources and procedures to investigate difficult questions. The organization can obtain and release outside assessments of potentially controversial internal evaluation reports. By taking such steps the organization strengthens the credibility of its internal evaluations.

17. Ensure that evaluations are fair and legally viable.

Organizations must design, conduct, and report evaluations that meet conditions of both fairness and legal viability (Zirkel, 1996). As much as possible, organizations should use evaluations for constructive purposes and steadfastly guard against destructive uses. They should assure that evaluation data are used for the authorized purposes. For example, when a school district or school gathers data exclusively to help a teacher diagnose and correct deficiencies in teaching, administrators should not subsequently use those same data to inform the public or fire or make other personnel decisions about the teacher. Organizations should define and implement guarantees of fairness to personnel in the collection and use of evaluation data.

However, an organization must not make promises to personnel that contravene pertinent organizational policies and laws. In some states the personnel evaluation files of all publically employed personnel by law are open to public review. In those states a public organization must candidly inform the staff that the courts might subpoena even formative evaluation data and make it available for public review. The general point here is that the organization's personnel must learn the policy and legal constraints on evaluations and within those constraints do everything possible to be fair to the evaluatees.

18. Ground evaluations in professional requirements for sound evaluation.

All service providers have an obligation to deliver high quality services and seek ways to improve the services. This applies every bit as much to their evaluations as to the other parts of their roles. An organization should set standards for its evaluation system. It should train all the staff to apply the standards. It should regularly use the standards to assess and improve its evaluation system.

The state of Hawaii exemplifies this recommendation. Its state board of education adopted professional standards for personnel and program evaluation as state policy. The state is now using the standards as the foundation for reforming its systems for assessing students, schools, the overall system, and administrators and teachers. It is using the standards as the basic

curriculum for training district personnel in the principles and procedures of program and personnel evaluation.

The preceding discussion of 18 checkpoints is intended to define the requirements of a sound organizational evaluation system. At first glance, the list may seem overwhelming. But it is in the best interests of an organization that seeks to succeed in serving clients to carefully consider every one of these checkpoints. The organization might decide to modify certain checkpoints to better fit its situation. But on close examination its leaders will find that every checkpoint is relevant to defining a sound vision for an organization's evaluation system.

Evaluation is complex. But it is also a fundamental component of every effective organization. Thus, it is important that the organization's leaders deal systematically with evaluation's complexities as they define their vision of sound evaluation. Only after they have articulated a sound vision can they decide on the components needed to implement that vision. Given the vision of effective evaluation seen in the foregoing discussion of the Type 1 Checklist, this article turns next to a discussion of what components are required to implement this vision.

Type 2 Checklist: Components of a Unified Evaluation System

The Type 2 Checklist appears in Exhibit 2. Its 10 checkpoints denote the operational components of a sound evaluation system. As with the Type 1 checkpoints, the Type 2 checkpoints are generic. They illustrate the types of components that any organization must put in place to carry out its evaluation program. But each organization should use these checkpoints as examples. The organization's leaders should adapt these checkpoints as appropriate to their situation.

1. Definition

The organization's personnel need to agree on a common definition of evaluation. In doing so they must avoid adopting any of the misleading or otherwise dysfunctional definitions of evaluation. For example, evaluation is more than measurement, more than judgment by an expert, and more than determining goal achievement (Stufflebeam et al., 1971). Moreover, it is not the same as empowerment or public relations (Stufflebeam & Webster, 1983; Stufflebeam, 1994). Nor is evaluation equatable to any specific procedure such as case study or experimental design (Joint Committee, 1994).

Nevertheless, the organization's personnel can choose from legitimate alternatives in defining evaluation. If the organization leaves open the question of definition, different members of the organization will likely work from different definitions of evaluation. The result will be confusion and lack of a unified evaluation approach. In concurrence with the North American Joint Committee on Standards for Educational Evaluation, the following definition is recommended:

Evaluation is the assessment of the merit and/or worth of some object.

This definition focuses on the root term in evaluation, which is value. It says that evaluations should assess the value dimensions called merit and worth. Merit concerns a thing's

EXHIBIT 2
TYPE 2 CHECKLIST:
COMPONENTS OF A UNIFIED EVALUATION MODEL

- 1. DEFINITION**
- 2. OBJECTS FOR EVALUATION**
- 3. STANDARDS**
- 4. COMMUNICATION MECHANISMS**
- 5. INFORMATION**
- 6. CRITERIA**
- 7. USERS & USES**
- 8. GENERAL FRAMEWORK**
- 9. SUPPORT STRUCTURE**
- 10. META-EVALUATION**

excellence. Worth concerns its cost-effectiveness in meeting clients' needs. Ideally, a particular program, project, service, or staff member should have excellent potential to serve the organization's clients. That is, it should have merit. Otherwise, it cannot serve clients well, and it may harm them.

But, merit is not enough. All sound services must also have worth. That is, they must address the clients' needs. For example, if almost no students need a particular university course, then the evaluation would indicate that it may not be worth offering, even if it would be excellently taught. In such cases, the university should consider terminating the course, or possibly strengthening its recruitment of students. Organizations waste resources when they sustain unneeded services, no matter their quality. Clearly, any service organization should assess both the merit and the worth of its various services.

But evaluators should be very careful in how they define and treat these concepts. The definitions of merit and worth are both context dependent. A squatter's shack in the city of San Carlos, Negros, the Philippines, has merit for those who would otherwise have no shelter at all and thus also has worth. In virtually any U.S. city, the same shack would be condemned and torn down, reflecting the judgment that it has neither merit nor worth. As someone has said, one man's treasure is another man's junk.

An advantage of this proposed definition is that it agrees with common dictionary definitions of evaluation and can be understood by personnel throughout an organization. Also, it is not technically complex.

2. The organization should clearly identify the objects that must be evaluated.

As noted above, an organization should evaluate all its aspects that collectively impact on its services to clients. To ensure that the organization's personnel pay more than lip service to these aspects, its leaders should explicitly identify the aspects to be evaluated and write them into the evaluation budget and schedule. For example, a school district might decide to evaluate the following aspects of its organization and programs:

- student background, needs, and achievements
- parent involvement
- district/school policies and rules
- governance and administration
- teacher competence and performance
- administrator competence and performance
- support services
- curricula
- extracurricular activities
- staff development
- school facilities and equipment
- budgets and fund raising
- district/school climate
- district-community relations

As this list illustrates, organizations must do a great deal of work in evaluating their capacity and contributions. They should have no “sacred cows” that are exempt from evaluation if they truly want to serve their clients well. In order to conduct the needed evaluations, an organization must involve all of its personnel, identify important issues to be assessed, assign priorities, realistically schedule and budget the evaluation work, and follow through.

Clearly, an organization does not need to evaluate everything every year. For example, it might reasonably evaluate its policies every three years. Nevertheless, over time the organization should evaluate and strengthen every organizational aspect that impacts on the extent and quality of its services.

If someone claims an exemption from evaluation because an aspect does not relate to organizational aims, then the organization should surely evaluate that aspect. The organization's leadership should probably jettison the aspect if it does not help in effectively serving clients or especially if it is counterproductive. In either case, the aspect would fail the worth criterion.

3. Adopt appropriate standards for use in guiding and assessing the organization's evaluation system.

Organizations must conduct principled evaluations. In establishing a defensible unified system of evaluation, an organization can take no more important step than to adopt appropriate evaluation standards. Such standards explicate the organization's definition of sound evaluation. They provide the organization's personnel with a common language about evaluation. More to the point, they provide organization-wide criteria and practical guidelines for planning, carrying out, and assessing evaluations. An organization can use the standards to set evaluation policies, to train the staff, and to provide the public and external groups a basis for assessing the organization's evaluations.

Fortunately, the evaluation field has developed professional standards in several areas of evaluation. In the U.S., there are professional standards for educational and psychological testing (APA, 1985), educational personnel evaluation (Joint Committee on Standards for Educational Evaluation, 1988), and program evaluation (Joint Committee on Standards for Educational Evaluation, 1994). All U.S. school districts and state education departments should consider adopting these three sets of standards as district or state policy as, for example, was done in Louisiana and Hawaii. Such standards provide the organizations with professionally endorsed evaluation principles and extensive practical guidelines. Also, experience and several studies have shown that the standards can be successfully applied to service organizations outside the field of education including the U.S. Marine Corps (Baker et al., 1995) and a division of General Motors (Orris, 1989), both of which adopted the Joint Committee personnel evaluation standards as a basis for assessing and improving evaluations of personnel.

The Joint Committee evaluation standards emphasize the imperative of stakeholder involvement in evaluation. Coalitions of more than a dozen U.S. and Canadian professional education organizations participated in developing these standards. They engaged teachers, administrators, lawyers, professors, school board members, educational researchers, curriculum experts, psychologists, psychometricians, counselors, and evaluators to collaborate in developing the standards. As a consequence, the standards are grounded in sound technical

requirements; are geared for use by teachers, administrators, support personnel, and school council members in schools; and reflect a broad consensus.

The standards are demanding but not esoteric. Teachers, administrators, and school board members had as much say in developing these standards as did statisticians and evaluation specialists. All the personnel of a school district, state education department, or other service organization can understand the plain language of these standards. These standards embody principles that have widespread applicability in evaluations: propriety, utility, feasibility, and accuracy.

The **Propriety Standards** require that evaluations be conducted legally, ethically, and with due regard for the welfare of clients and stakeholders. These standards address issues concerned with effective service to clients, formal evaluation guidelines, conflict of interest, rights of clients and evaluatees, access to evaluation reports, and professionalism and sensitivity in interacting with evaluatees.

The **Utility Standards** require evaluators to provide information for use in improvement and accountability. The organization must ground its evaluations in appropriate values and criteria. The evaluations should address issues of importance in improving the organization. The organization must train staff to conduct sound evaluations so that their reports will be valid, trusted, and acted upon. The evaluators must issue timely, relevant reports. And the organization's personnel must use the findings to strengthen their services and to earn and maintain public credibility. The utility standards require that organizations allocate their evaluation resources only to worthwhile studies. The staff should conduct, report, and use the studies so that they will make a positive difference in the organization and in their individual contributions. According to the utility standards, nothing is much worse in evaluation than an academic study that has no beneficial use.

The third subset of the Joint Committee Standards focus on issues of feasibility. The **Feasibility Standards** require organizations to conduct evaluations that are realistic, prudent, diplomatic, and frugal. As much as possible the organization should employ practical evaluation procedures that fit into its overall routine. It should make appropriate use of machine scorable evaluation instruments. It should not jump on any bandwagon to discard objective measures, such as multiple choice testing, as these can provide great scope and high reliability in assessments at low cost. It should consider embedding evaluation into its regular programming. It should encourage and support every staff member to integrate evaluation into her or his work. The organization should also make its evaluations cost-effective. It should conduct only those studies that are needed. It should expend only the amount of personnel time and money required to meet an evaluation's purposes. Organizations should make appropriate use of sampling when they do not need data from every potential respondent. Organizations should carefully examine their data collection systems and plans to assure that they need and will use all the involved data. They should not collect data if they won't use it. Finally, the feasibility standards require that organizations establish and sustain a participatory approach to evaluation. For example, school districts should involve teachers, parents, support staff, school board/council members, students, and other stakeholders in planning, conducting, and using the evaluation results. Such a participatory approach helps to get the evaluation work done. It also prepares organizational stakeholders to understand, accept, and act upon the evaluation findings. Probably, most service organizations would readily adopt the Joint

Committee's subset of standards requiring organizations to keep their evaluations within reasonable bounds of feasibility.

The Joint Committee's fourth and last subset of standards are the **Accuracy Standards**. According to these standards, organizations must ground their evaluations in sound information. They must clearly identify the object of the evaluation and the study's procedures. They must take into account the organization's context and the characteristics of its clients. They must gather information that is both reliable and valid. They must assess and control bias. They must issue justified conclusions. And they must monitor, evaluate, and improve all aspects of the evaluation system.

Clearly, an organization can benefit by adopting and applying appropriate evaluation standards. By meeting such standards, an organization assures that its evaluations will be ethical, useful, feasible, and accurate.

4. Organizations should install mechanisms to effect needed communication in evaluations.

Sound evaluations require effective communication every bit as much as they require sound data collection procedures. The need for communication is pervasive in the evaluation system and in every specific study. The organization must clarify its evaluation system to its personnel and constituents. Members of specific evaluation teams must make sure that they and their audiences agree on the purposes of the evaluation. The evaluation team must clearly report its findings, so that users make defensible inferences. In other words, effective evaluation necessarily involves much interaction and negotiation.

An organization may employ a number of concrete steps to meet needs for effective communication in its evaluation work. The following nine such steps have been used effectively in school districts and other organizations.

- The organization can **use focus groups and task groups to involve stakeholders in designing its evaluation system**. Hawaii has done this very effectively in its process of assessing and redesigning its systems for assessing students, schools, administrators, teachers, support personnel, and the overall state system.
- Once the evaluation system has been designed, the organization's leaders should **develop and disseminate a printed description of the system**. This will help the organization's personnel and constituents to develop a shared understanding of the official evaluation approach.
- Evaluation teams can **conduct public meetings and meet with important leadership groups to report evaluation findings and learn audience reactions**. In arranging such meetings, the evaluation team should consider preparing a press release and providing advance copies to representatives of the public media. The Hawaii Department of Education has regularly informed the state's senate, house, and state board of education about the state's standards for evaluations and its progress in developing a comprehensive assessment and accountability system, and this has promoted understanding and support for the work.

- To strengthen involvement of constituents in evaluating organizational functions, an organization's leaders could **appoint and meet regularly with a standing, representative accountability commission**. The organization's leaders might choose commission members from clients, the public media, the business community, the clergy, the juvenile court, higher education, local schools, local law enforcement, area philanthropic organizations, and local government. The organizations leaders should engage this commission to help set priorities for future evaluation studies, define evaluation questions of interest to sponsors and constituents, and propose criteria for evaluating particular organizational functions. The organization's leaders could also engage the commission to critique draft evaluation reports, discuss findings, and help communicate the findings to the broader audience. The organization's leaders might also ask the commission to help raise funds or other resources to assist the evaluation efforts. The Dallas Independent School District exemplifies this recommendation through its engagement of a standing, broadly representative accountability commission (Webster, 1994).
- Evaluators should **identify and involve their clients**. Every evaluation includes a client as well as an evaluator. Sometimes they are the same person, as in a teacher's self-assessment. But usually there is a separate client. In such cases the evaluator should identify the client and other representatives of the evaluation's audience. The evaluator should engage these persons to help clarify the most important evaluation questions, criteria, and audiences for reports. This will help the evaluators to address questions of importance to the audiences. The exchange will also prepare the client and other participants to understand, accept, and use the evaluation findings.
- **Issue clear, brief, pointed reports**. The evaluator should deliver reports that the intended audience can and will read. The main report should immediately engage the audience in reading its contents. It should provide a minimum of background information, then focus on the evaluation questions. It should directly present the main findings. It should be straightforward and easy to grasp. It might employ headlines, pictures, and graphs to emphasize the main points. It might use vignettes to illustrate the main findings. It should tell the reader how to get the evaluation study's technical report. The evaluator might prepare multiple versions of the main report to serve the needs of different audiences. For example, an organization might issue both Spanish language and English reports or summary and full-length reports. In addition, the evaluator might develop audiovisual materials to assist in presenting the written reports.
- **Deliver the reports when they are needed**. Evaluation reports lose much of their potential utility when delivered late. When planning an evaluation, the evaluation team should forecast times when the client will need evaluation feedback. The team should plan its data collection and reporting activities accordingly. During the evaluation process, the team should interact regularly with the evaluation's users to remain current about the timing of their needs for evaluation feedback and opportunistically to inject useful feedback into the decision process. When feasible, the evaluators should adjust their evaluation timetable to maximize the evaluation's utility to clients. Among the skilled evaluators who have exemplified this recommendation are Howard Merriman when he headed the Columbus, Ohio, school district's department of evaluation and William Webster, director of evaluation in the Dallas Independent School District. Both

of these men regularly attended school board and administrative cabinet meetings as a vital means of projecting needs for evaluative information.

- **Provide feedback workshops to clients** to go over draft evaluation reports. This practice has a number of important advantages. It provides a means of early feedback before the evaluation report is finalized. It engages the client to help identify any ambiguities and inaccuracies in the draft report. It provides the client an opportunity to update the evaluator on recent developments. It prepares the client group to receive and use the finalized report. And it provides a venue for determining how best to disseminate the final report.
 - However, client/evaluator review of the draft report can be problematic. The evaluator must resist attempts by the client to remove or gloss over unwelcome, though valid evaluation findings. In such cases, the evaluator should accept input, promise to consider it, but emphasize that he or she must finalize the report in accordance with the appropriate evaluation standards.
 - Communication is such a vital part of the evaluation process that the organization should **consider** periodically **engaging a communications expert to help improve the communication aspects of the evaluation system**. The organization could ask the expert to assess and provide direction for strengthening all the communication aspects of evaluation discussed above.
5. **The organization should collect and use a broad range of information to evaluate its services.**

Writers have proposed various checklists and lists of variables to help organizations identify the needed information. These writers advise evaluators to consider more than client outcomes when evaluating services. In previous writings (Stufflebeam, 1983, 1988) I have proposed that evaluators collect four generic types of information. These deal with the school district's context, inputs, processes, and products. Using this classification system, evaluators at any level of an organization can identify virtually all the information needed in broad-gauged evaluations.

Context information illuminates the pertinent environment of, for example, a classroom, school, or district. It includes client needs, organizational strengths, outside opportunities for strengthening the organization or a part of the organization, problems to be solved, and goals and priorities that might require revision. This information is important for driving and assessing improvement efforts, including the organization's strategic planning process. It helps members of the organization to set goals that reflect assessed needs. It identifies shortcomings and problems in the organization or part of the organization under review. It also denotes assets that the organization or individual staff members can use to solve problems and better meet client needs. Context information also provides the most important criteria for judging the success of improvement efforts. These criteria are the clients' assessed needs. Organizations can judge programs and services successful only when they help meet the needs of targeted clients.

Input information includes assessments of the organization's present strategies, resource allocations, and work plans for serving clients and assessments of alternatives that might be adopted. Pertinent input variables include policies, the organization's planning process, budgets, fund-raising plans, calendars and schedules, staffing plans and assignments, staff development plans, approaches to community involvement, arrangements for maintaining facilities and equipment, and project designs. Other input variables are products and services under consideration for purchase. In a school district these might include textbooks and other instructional materials, testing programs, and instructional equipment. Organizations should conduct input evaluations before funding new services or revising existing services.

Process information includes descriptions and judgments of the efforts to deliver services. In examining process, the evaluator describes and judges staff activities, records and analyzes costs, and diagnoses flaws in service delivery plans. The organization's leaders and staff members need process information to guide and monitor the delivery of services and discharge of other organizational functions. Also, evaluators must record and analyze process to help interpret why a project succeeded or failed.

Product information includes descriptions and judgments of outcomes. In school districts these especially include assessments of student achievements. Product evaluations identify and examine both intended and unintended outcomes. They compare the outcomes to assessed needs of the clients. They also consider whether any shortfalls are more a function of a faulty program plan or poor implementation. Product evaluations may look definitively at **impact** or the extent to which services reached the intended beneficiaries, **effectiveness** or the quality of impacts, and **viability** or the sustainability and transportability of the enterprise under review.

Exhibit 3 summarizes the indicators that apply to context, input, process, and product information in a school district. Exhibit 4 summarizes the indicators that Western Michigan University is using to help the Alger Foundation to evaluate its long-term success in supporting and operating community development programs in the Philippines and Hawaii.

6. Organizations should employ multiple levels of criteria in evaluating services.

These include basic societal values; criteria fundamental to any evaluation; criteria associated with the organization's context, input, process, and product; criteria reflecting the generic duties of the organization's personnel; and ground-level criteria that inhere in particular services or programs. Exhibit 5 presents the main features of these five sets of criteria.

As denoted in the chart, the organization should ground its evaluations in basic societal values. Over the past 30 years, the U. S. has experienced an evolution of such values.

In the 1960s U.S. leaders finally noticed that the public schools and other service organizations were providing not only segregated but also low quality service to low income persons,

EXHIBIT 3
GENERIC TYPES OF EVALUATION INFORMATION

CONTEXT

(NEEDS, OPORTUNITIES, PROBLEMS, GOALS)

INPUTS

(PLANS, RESOURCES, PERSONNEL, SCHEDULE)

PROCESS

(IMPLEMENTATION, COSTS, DESIGN FLAWS)

PRODUCTS

(OUTCOMES, SIDE EFFECTS, IMPACTS)

EVALUATION INDICATORS

CONTEXT EVALUATION

<p>Quality of Life: Livelihood Health Safety</p>	<p>Nutrition Housing Social</p>	<p>Sanitation Water quality Family structure</p>	<p>Family values Family relations Living arrangements Education/training</p>	<p>Community Settings: Resource organizations Government services Recreation opportunities</p>	<p>Political climate employment Economy</p>	<p>Related programs Private sector leaders</p>
---	---	--	--	---	---	--



PRODUCT EVALUATION

Impact Evaluation
Directory of beneficiaries
Percent of target group served
Service to hard-to-reach beneficiaries
Other groups served
Services to the wrong group
Levels of participation
Dropouts
Empowerment of target audience
Impacts on integration
Realignment of power
Effects on the community
Redefinition of target group as appropriate
Adaption of service strategy as appropriate
Incidents of corruption

Effectiveness Evaluation
Full range of outcomes
Depth of effects
Quality of effects
Significance of effects
Quality of life changes
Short-term outcomes
Long-term outcomes
Unintended outcomes
Cost-effectiveness
Continuing improvements
Diagnosis of failures
Awards
Attitude changes
Values changes

Viability Evaluation Sustainability
Institutionalization
Plans and actions
Demand for continued service
Area support
Area constraints
Financial commitments
Legal arrangements
Financial viability
Legal viability
Staffing viability
Threats
Transportability:
Potential adopters
Actual adoptions
Inquiries
Visitors



PROCESS EVALUATION

Supervision:
Staff orientation & training
Scheduling
Directing, controlling, assessing
Communicating with staff
Implementing plans
Quality & timing of task performance
Progress toward objective

Resource management:
Purchase of goods & services
Fiscal records
Safeguards against graft
Resource utilization
Cost overruns
Cost overruns
Maintenance of security

Quality control:
Inspections
Internal evaluation
Progress reports
Workability of plans
Review & improvement of plans
Correction of operational problems
Other uses of evaluation
Record of actual progress

External relations:
Participation by collaborators
Effective public relations



INPUT EVALUATION

Planning:
Values clarification
Focus on priority needs
Background discussions
Review of research
Study of similar projects
Assets analysis
Political analysis
Legal analysis
Defined target group
Clear goals
Sound project strategy
Detailed plan
Comprehensive, coherent, pertinent, superior plan
Paradigm shifts

Preparations:
Policy decisions
Commitment of resources
Budget
Political arrangements
Provisions of orienting, training, and evaluating staff
Work schedule
Chain of command
Other protocols
Contracts & subcontracts
Memoranda of agreement
Facilities
Equipment
Safety standards
Arrangements for evaluation
Accounting mechanisms
Arrangements for collaboration
Publicity



EXHIBIT 4

EXHIBIT 5

EVALUATION CRITERIA

Basic Societal Values

- Equality of Opportunity
- Effectiveness
- Feasibility
- Excellence

Generic Evaluative Criteria

- Merit
- Worth

CIPP Criteria

- Client Needs
- Responsiveness and Merit of Work Plans
- Congruence Between Activities and Plans
- Quality and Significance of Outcomes

Generic Duties

- Line Staff Duties
- Support Staff Duties
- Administrator Duties
- Board Duties

Ground-Level Criteria That Requires
Negotiation

especially poor black children and adults. Subsequently, the federal government enacted strong laws designed to guarantee **equality of opportunity** in the publicly supported institutions. Equality of opportunity is now an entrenched value in American society, but not necessarily a reality in every public service organization. Every U.S. public school district is required by law to deliver equitable educational opportunities to all its students and is thus subject to governmental inspection on this criterion. U.S. public schools and other publically supported service organizations must be prepared to report on their fulfillment of the equal opportunity criterion.

Also, in the 1960s and 1970s, the U.S. society invested heavily in improving schools and other service organizations. Hubert Humphrey, John Kennedy, Lyndon Johnson, and other leaders concluded that the equal opportunity movement would not succeed without extensive improvement of the schools and other service organizations. As a part of its War on Poverty, the U.S. Congress invested billions of dollars to improve the schools and other service organizations. Along with this huge infusion of funds, the federal government required these organizations to be accountable for using the funds effectively. This requirement ushered in the age of accountability and the basic value of **effectiveness**. For the first time in U.S. history, the federal government required the publicly supported organizations to be accountable for using federal funds effectively to serve clients, especially members of minority groups. Throughout the 1970s the U.S. government poured unprecedented levels of federal money into schools and other service organizations. Many of these organizations got accustomed to money for consultants, additional staff, special services, new materials, evaluation staffs, travel to conventions, modern equipment, school reorganization, etc.

In the early 1980s reality struck the service organizations, including schools, universities, and welfare agencies. The U.S. economy went into a period of serious decline. The government was continuing to invest huge sums of money in the service organizations, but accountability reports from these organizations showed little evidence of success in providing improved services to clients. Government officials and many citizens became disillusioned with the apparent failures of funded organizations to improve achievement and employment among the poor. On into the 1980s, both the federal and many state governments cut back funding of public services. Consequently, many service organizations had to reduce staff, discontinue programs and, in general, cut services. These organizations were abruptly introduced to a new societal value: **feasibility**. In evaluating plans and operations, service organizations were directed to seek efficiencies, stay within their means, and emphasize cost-effectiveness.

The 1980s saw society usher in yet a fourth value to be applied, especially to the public schools and the auto industry. W. Edwards Deming (1982, 1986) and others convinced society that America's economic downturn was largely due to a decline in international competitiveness. Deming said this was due to poor quality of American products. Some pundits went the next step to argue that erosion of quality in American products was due to too much emphasis on equity and affirmative action and not enough on quality. Tom Peters (1982) highlighted the concern for quality in his book *In Search of Excellence*. Subsequently, the auto industry and other organizations throughout American society undertook steps to improve the quality or merit of their products and services. School districts were no exception. Among the schools' moves toward excellence are Total Quality Management, quality circles, magnet schools, and charter schools and other schools of choice.

The four societal values discussed above are now crucially important in evaluating U.S. service organizations. These are equality of opportunity, effectiveness in serving clients, feasibility or prudence in using resources, and excellence of service and outcomes. The key points are that organizations should include basic societal values in evaluating programs and services and that such values evolve over time. Clearly, organizations should regularly engage in values clarification.

The second set of evaluation criteria included with Checkpoint 6 derive from the proposed definition of evaluation. As discussed previously, evaluations necessarily assess **merit and worth**. Merit concerns a thing's intrinsic value or quality. It is akin to the societal value of excellence described above. Worth involves a thing's extrinsic value, or how useful it is in meeting assessed needs. According to the proposed definition of evaluation, organizations should evaluate the merit and worth of their services. Merit is the first-order criterion. A thing can have merit without worth. But it cannot be worth much if it has no merit.

The Context, Input, Process, and Product categories of information discussed above suggest a third set of evaluative criteria. The most important of these are assessed **client needs; quality and feasibility of plans; responsiveness of plans to assessed needs; congruence between activities and plans; and range, quality, significance, and cost-effectiveness of outcomes**.

A fourth set of criteria are especially important in evaluating the performance of personnel. The organization determines these criteria by examining the **duties** assigned to each staff member and to groups of personnel such as special committees and the organization's policy board. For example, a teacher's duties might include knowledge of content, classroom management, effective communication of content to students, assessment of student needs and achievement, fostering parent involvement, counseling and referring students, and cooperating in school improvement efforts. The school leader's duties might include promoting and supporting student growth and development, promoting equality of opportunity, fostering a positive school climate for learning, leading school improvement efforts, strengthening classroom instruction, managing personnel, managing the district/school's finances and facilities, assuring school safety, effecting positive school-community relations, fostering staff development, and collaborating with the school board/council. Many service organizations invest most of their resources in personnel. Therefore, personnel evaluation is crucially important in improving the organizations. An organization should clarify each member's duties. These duties provide the most relevant criteria for assessing and strengthening performance of the personnel.

This discussion suggests that the matter of evaluative criteria is extremely complex. So far, the discussion has included basic societal values, criteria inherent in the definition of evaluation, criteria inherent in the types of information to be collected, and criteria associated with staff duties.

There is yet a fifth level of criteria. These cannot be prespecified. Michael Scriven calls these the **ground-level criteria**. They are idiosyncratic to particular evaluations. An organization's staff must conceptualize and negotiate these specific criteria when planning a particular study. One can do this by studying relevant background information, holding a discussion with the client, and conducting focus group meetings with stakeholders to help clarify the key issues. One might also study reports from past evaluations of similar services. Moreover, some of

these criteria may not be clear until the evaluation is well under way. Again, organizations should be flexible in designing and conducting evaluations, so they can continually improve the evaluation criteria and data collection plan. The main points here are that one can never predetermine all the criteria needed in a given evaluation and that these must be determined in concert with stakeholders. One must work very hard and thoughtfully throughout the evaluation process to derive and apply the appropriate criteria.

7. Clearly identify the intended users and uses of the organization's evaluations.

This checkpoint directs the organization (or staff member or team) to make an analysis of who will use its evaluation findings and how they will use them. The evaluator needs this information to tailor evaluation reports to the user's questions and information needs. Exhibit 6 illustrates such an audience analysis. According to this analysis, an organization's evaluation system serves two primary groups. The first includes the organization's personnel. In school districts, these are the teachers, administrators, support personnel, school board/council, and students. The second group includes persons outside the organization. They might include taxpayers, accrediting organizations, government agencies, colleges and universities, and employers. Thus, an organization's evaluation system or an individual evaluation may well need to serve a diverse audience.

While members of the audience will surely vary in their interests in the organization, program, or service being evaluated and their related information requirements, they are likely to use evaluation reports in one or two main ways. These are for formative or summative purposes.

Formative evaluations provide ongoing feedback to help the organization's personnel manage and improve services. Staff members conduct ongoing formative evaluations. Through this process, the organization's personnel identify client needs and assess plans, operations, and outcomes. They use the findings to strengthen services, programs, projects, and the performance of staff members. They also use formative data to develop and release progress reports. In this sense, formative evaluations provide early warnings of what the final summative evaluation results will be if the present course is continued. External evaluators and other outside audiences often are keenly interested in finding out whether an organization's staff conducts and uses formative evaluations to improve its programs and services.

Summative evaluations are conducted after the completion of a service cycle or program. Often, external evaluators conduct the summative evaluations of programs. The organization's leaders conduct the summative evaluations of staff. The evaluators usually conduct summative evaluations at the end of the year, at the end of a two- or three-year project period, or at the end of a staff member's probationary period. They design summative evaluations to inform the organization's leaders and sometimes its external constituents of the merit and worth of the organization, some more specific part of the organization, or the performance of a staff member. In the case of staff evaluations, organizational officials use summative evaluations to inform decisions on assignment, promotion, tenure, merit pay, special awards, or firing. Policy and funding groups use summative evaluations to decide whether to retain, alter, or discontinue a program or service. The public uses summative evaluations to inform its

EXHIBIT 6

USERS & USES OF EVALUATION

USERS

USES

ORGANIZATIONAL PERSONNEL

FORMATIVE (IMPROVEMENT)

SUMMATIVE (ACCOUNTABILITY)

- LINE STAFF

X

X

- ADMINISTRATORS

X

X

- SUPPORT

X

X

- BOARD

X

X

- CLIENTS

X

OTHER CLIENTS

- FUNDERS

X

- TAXPAYERS

X

-ACCREDITING

X

- GOVERNMENT

X

- LOCAL

X

- MEDIA

X

opinions of an organization or more specific service. In general, organizations conduct or commission summative evaluations to make important funding and personnel decisions, meet external accountability requirements, and maintain credibility in the community.

Bob Stake provided a simple example to distinguish between formative and summative evaluation. He said “When the cook tastes the soup, that’s formative. When the guests taste the soup, that’s summative.” Michael Scriven emphasized that formative and summative evaluations differ little in their use of criteria and information. He said each formative evaluation is a preliminary version of the final summative evaluation. Thus, formative evaluations provide decision makers with opportunities to correct mistakes and improve a service before the final judgments. Summative evaluations provide no such in-process opportunities for improvement. They lay it on the line concerning the object’s merit and worth. It is then up to the public and other constituents to draw conclusions and make appropriate choices.

In summary, Exhibit 6 shows that an organization’s personnel need both formative and summative evaluations. For the most part, external audiences need summative reports. Parents are the lone exception in the external group. They need ongoing formative evaluation of their child’s achievement to help guide the learning process. Of course, with the advent of charter schools and other choices available to parents, they need summative evaluations of schools, too, so they can pick the one that is best for their child.

8. The organization should adopt a general conceptual framework.

This paper employs two checklists to help an organization’s personnel see and understand the complexities of systematic evaluation. An organization’s leaders also need simplified devices to help staff and constituents remember, discuss, and apply the core concepts in the organization’s evaluation approach. Checkpoint 8 is intended to address this purpose. Exhibit 7 contains a general framework for planning evaluation studies. This two-by-four matrix relates four types of evaluation to two main uses of evaluation. The four types are labeled context, input, process, and product evaluation, corresponding to the four general types of information needed to evaluate school district services. The two uses are formative and summative evaluation, as discussed above. As seen in Exhibit 7, this matrix identifies eight types of evaluation studies.

Formative context evaluations provide guidance for setting objectives and priorities. Formative input evaluations provide guidance for planning programs and services. Formative process evaluations provide guidance for delivering services. Formative product evaluations provide guidance for recycling, continuing, or improving services. All four types of formative evaluation focus on helping an organization’s personnel to improve their services.

Summative evaluation, on the other hand, provides an organization’s leaders and outside audiences with information for judging completed efforts or summing up the value of some service, based on its track record. Summative context evaluation provides information for judging whether objectives and priorities are consonant with assessed client needs. Summative input evaluations assess whether strategies are responsive to assessed needs and superior to alternatives. Summative process evaluations help audiences assess the extent to which plans were well implemented. Summative product evaluations help audiences identify

EXHIBIT 7
 GENERAL FRAMEWORK FOR EVALUATION:
The Main Contributions

**TYPES OF
 EVALUATION**

USES OF EVALUATION

	FORMATIVE	SUMMATIVE
CONTEXT	Guidance for setting objectives and priorities	Comparison of objectives to needs
INPUT	Guidance for planning programs and other services	Comparison of plans to alternatives
PROCESS	Guidance for delivery of services and implementations of programs	Record of implementation
PRODUCT	Guidance for recycling/continuation decisions	Comparison of achievements to needs, objectives, and priorities

outcomes and assess their significance compared to assessed needs, organizational objectives and priorities, and the adequacy of the process.

School districts often conduct the four types of formative evaluation separately and more or less sequentially. This is because formative evaluations are proactive. They track and guide the development and implementation of projects or other efforts. This is portrayed in Exhibit 8. It shows that a school district might conduct context, input, process, and product evaluations sequentially approximately in different quarters of the year. The district might conduct context evaluation in the first quarter to help set district objectives and priorities. It could follow in the next quarter with input evaluation to assist in finalizing the district's annual plan. The district's evaluators could then apply process evaluation in the third quarter to guide and record the staff's implementation of the district plan. The evaluators would then conduct product evaluation in the fourth quarter to identify and assess outcomes. The evaluators would next feed the product information into the next year's context evaluation.

In contrast to the prospective nature of formative evaluations, summative evaluations are retrospective. They look back on what was done and accomplished. Thus, the four types of summative evaluation often are conducted simultaneously. Ultimately, the findings from summative context, input, process, and product evaluation are synthesized into an overall assessment of the merit and worth of the organization, program, or particular service.

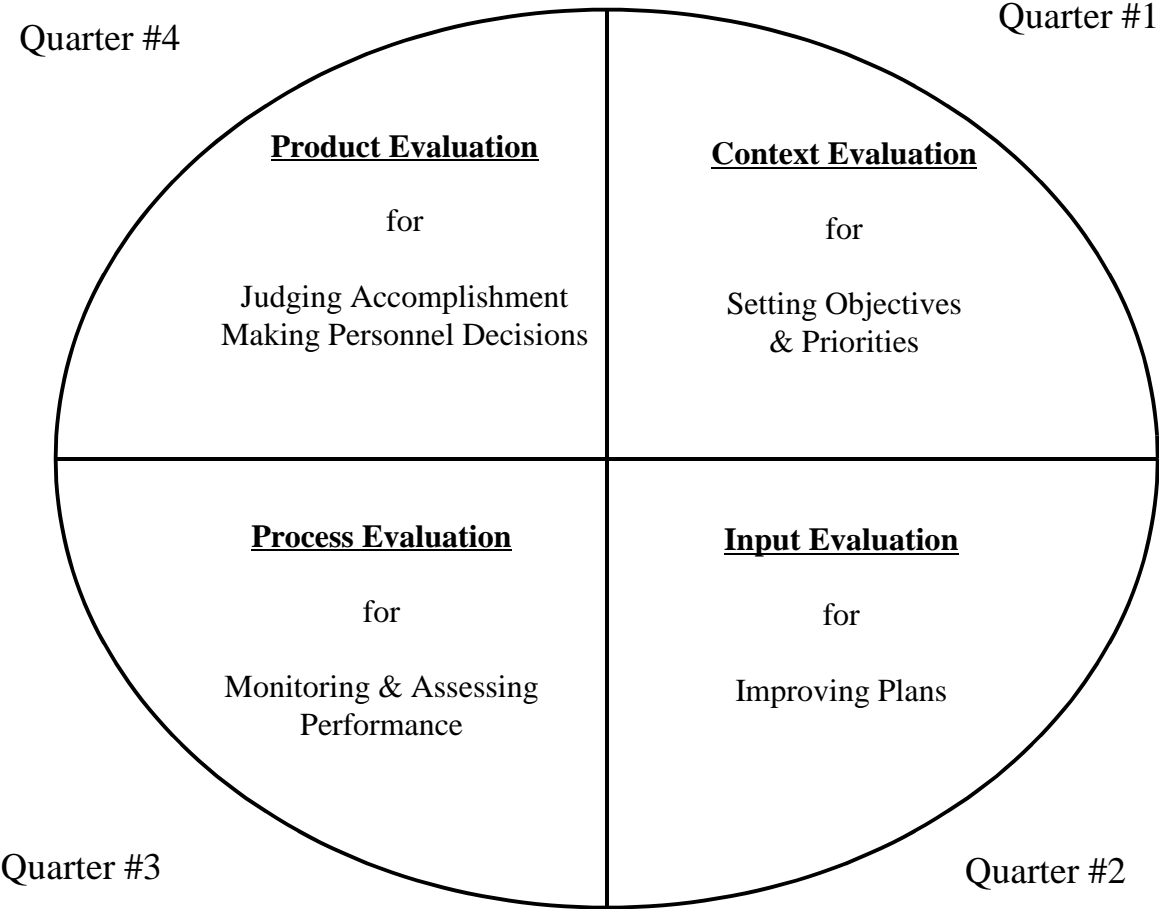
Arguments for using the framework presented in Exhibit 7 are that it is comprehensive, easy to remember, and useful for guiding decision making and generating accountability reports. Also, it promotes efficiency in data collection. Formative evaluations provide data needed to manage and improve an organization's efforts. The organization can use these same data to develop the needed accountability reports for use in summative evaluations.

9. The organization should provide a strong support structure for evaluation.

Whatever evaluation approach an organization chooses, it must provide the resources and other support required to carry out the needed evaluations. To implement the evaluation approach described in this article, the organization should implement steps such as the following:

- Encourage and provide **time and support for self-assessment**.
- Provide **inservice education in evaluation**.
- Support the organization's staff to participate in **evaluation conferences**.
- Maintain up-to-date **evaluation policies** that clarify the staff's evaluation responsibilities and the organization's support for evaluation work.
- Maintain a **library of evaluation literature, instruments, and other resources**.
- Maintain **databases** for use in the organization's various evaluations.

Exhibit 8
MAIN EVALUATION EMPHASIS
in Each Quarter



- Provide support for **evaluation consultants**.
- Consider employing one or more **evaluation specialists**.

Evaluation is a critical organizational function. It is essential to an organization's success. It should be done thoroughly, usefully, and well. It should occur at all levels of the organization. This will happen only if the organization allocates a high priority to evaluation. The organization must manifest such a priority by committing the needed time and resources. None of the preceding suggestions in this realm is unreasonable.

10. Organizations should evaluate their evaluation systems.

The thesis of this paper is that an organization should regularly evaluate and strengthen every organizational function that impacts on client services. Evaluation is one of the most important such functions. The organization should regularly evaluate its evaluation system both to improve it and to demonstrate that its evaluations are sound and cost-effective.

The label metaevaluation is typically applied to evaluations of evaluations. Metaevaluations are critically important to the success and credibility of service organizations. They guard against evaluations that might mislead decision makers or gloss over serious problems. They are essential for instilling public confidence in the organization's evaluation reports. They are also needed to assure that their investments in evaluation are sufficiently helpful in improving services to warrant costs. For an example of a comprehensive metaevaluation report, see Finn, Stevens, Stufflebeam, and Walberg (1997).

An organization can meet the metaevaluation requirement by carrying out steps such as the following:

- Regularly **apply professional standards** of evaluation and measurement to plan and assess the organization's evaluations.
- Encourage and support staff to conduct **internal metaevaluations** as a means to both quality assurance and learning from their evaluation experiences.
- Consider engaging an **evaluation commission** to advise staff on evaluation policy and evaluate the organization's evaluation plans and draft reports.
- Periodically engage an **external metaevaluator** to evaluate and provide recommendations for strengthening the organization's evaluations.
- Employ metaevaluation results to **improve propriety, utility, feasibility, and accuracy** of evaluations.
- Periodically **reference the Type 1 checklist to assess the organization's progress in achieving the goals of a unified evaluation system**.

Summary and Conclusions

This article is based largely on an article of faith. It is that service organizations are comprised of professionals. Such professionals must constantly seek to effectively serve all their clients. They must have high standards of service. They must examine their practices against the standards. They must seek to improve wherever their work is deficient and wherever the state of the art has validated better strategies to serve clients. They must work collaboratively with colleagues, clients, and community to assess and effectively address the full range of needs of prospective and actual clients. They must earn credibility with clients, regulatory groups, funders, and the community. They must be open to and interested in having their work evaluated by others. But they must not await or be totally dependent on having outsiders evaluate and direct improvements in their work. Instead, the effective and socially responsible organization is one whose professionals regularly conduct and use evaluations to improve services to clients and inform constituents about the organization's accomplishments, failures, and needs.

Under the foregoing supposition, this article provides service organizations with conceptual and practical advice for strengthening their evaluation systems. It argues that every service organization needs to conduct sound evaluations to effectively serve clients, plan and effect improvements in its services, and earn credibility with clients, sponsors, and others. Moreover, an organization can best meet its evaluation needs by employing a sound unified evaluation system. In fact, any organization that practices participatory management must be guided by a unified approach to evaluation.

Under a unified evaluation system, an organization's evaluations of clients, services, special programs, and personnel

- follow the **same general model**
- adhere to the **same ethical and technical standards**
- promote **excellent and equitable service to clients**
- **involve the organization's personnel collaboratively** in using evaluations to strengthen operations and services and maintain accountability

To establish a fully functional unified evaluation system, organizations need to evaluate and strengthen their present evaluation practices. This article is built around 2 checklists designed to facilitate evaluation improvement projects. Basically, these checklists are planning tools. They present illustrations of what is included in a sound unified approach to evaluation. The checklists are not prescriptions for every service organization. The Type 1 Checklist suggests 18 features or aims of a sound unified evaluation system. The Type 2 Checklist presents 10 components required to achieve the 18 aims. Different organizations might choose different goals for their evaluation systems. And an organization's particular goals for evaluation may require components that differ from those recommended in this article.

The crucial point is that an organization periodically should examine and strengthen its evaluation system. It is only through sound evaluation that the organization's staff can determine strengths and weaknesses in its offerings. Such determinations are essential for deciding in what respects services should be improved. Continual improvement is the hallmark of a truly professional service organization.

This article was written partially to convince leaders of service organizations that sound evaluations are fundamentally important in improving services. In addition, the included checklists are intended to assist organizations to think through and address the issues involved in developing and applying a fully functional, unified approach to evaluation.

Bibliography

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological tests*. Washington, DC: Author.

Baker, W., Hattie, J., Horn, R., Jaeger, R. M., & Stufflebeam, D. L. (1995). *Evaluation of the performance evaluation system: Task 1 report*. Kalamazoo, MI: The Western Michigan University Evaluation Center. (Note: All reports for the U.S. Marine Corps project are distributed by Richard Voltz, MCSS Management Branch, U.S. Marine Corps, 3300 Russell Road, Quantico, VA 22134-5130.)

Deming, W. E. (1982, 1986). *Out of the crisis*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Finn, C. E., Stevens, F. I., Stufflebeam, D. L., & Walberg, H. J. (1997). A meta-evaluation. Chapter VI in *The New York City public schools integrated learning systems project. International Journal of Educational Research*.

Guba, E. G., & Stufflebeam, D. L. (1970, June). *Strategies for the institutionalization of the CIPP evaluation model*. Address delivered at the 11th Phi Delta Kappa symposium on Educational Research, Columbus, OH.

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.

Orris, M. J. (1989). *Industrial applicability of the Joint Committee's personnel evaluation standards*. Unpublished doctoral dissertation. Kalamazoo, MI: Western Michigan University.

Peters, T. (1982). *In search of excellence: Lessons from America's best-run companies*. New York: Harper & Row.

Sanders, W. L. & Horn, S. P. (1993). *The Tennessee value-added assessment system (TVAAS): Mixed model methodology in educational assessment*.

Stufflebeam, D. L. (1983). The CIPP Model for program evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models*. Boston: Kluwer-Nijhoff.

Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, 15(3), 321-338.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.

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

Stufflebeam, D. L., & Shinkfield, A. J. (1984). *Systematic evaluation: A self-instructional guide to theory and practice*. Boston: Kluwer-Nijhoff.

Stufflebeam, D. L., & Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models*. Boston: Kluwer-Nijhoff Publishing.

Stufflebeam, D. L., & Webster, W. J. (1988). Evaluation as an administrative function. In N. J. Boyan (Ed.), *Handbook of research on educational administration*. White Plains, NY: Longman.

Tymms, P. (1995). Setting up a national "value-added" system for primary education in England: Problems and possibilities. Presented at the CREATE National Evaluation Institute, Kalamazoo, MI.

Webster, W. J. (1994). The connection between personnel evaluation and school evaluation. In A. McConney (Ed.), *Toward a unified model: The foundations of educational personnel evaluation*. Kalamazoo, MI: The Western Michigan University Evaluation Center.

Webster, W. J., Mendro, R. L., & Almaguer, T. D. (1993). *Effectiveness indices: The major component of an equitable accountability system*. ERIC TM 019 193.

Webster, W. J., Mendro, R. L., & Almaguer, T. D. (1994). Effectiveness indices: A "value added" approach to measuring school effect. *Studies in Educational Evaluation*, 20, 113-145.

Zirkel, P. A. (1996). *The law of teacher evaluation: A self-assessment handbook*. Phi Delta Kappa Educational Foundation in cooperation with the National Organization on Legal Problems of Education, Bloomington, IN.