Program Evaluation in the GPRA Environment

GPRA seeks to promote a focus on program results by requiring agencies to set program and agency performance goals and report annually on their progress in achieving these goals. GPRA recognizes the complementary nature of program evaluation and performance measurement. Both are important components of an effective performance measurement system.

Pre-GPRA

Before the enactment of GPRA, our agency was conducting numerous program evaluation studies. For example, Congress had mandated comprehensive National assessments of both the Title I and Chapter 1 programs. When we began to develop performance plans under GPRA, we included these activities as part of our strategy for capturing performance data.

Post-GPRA

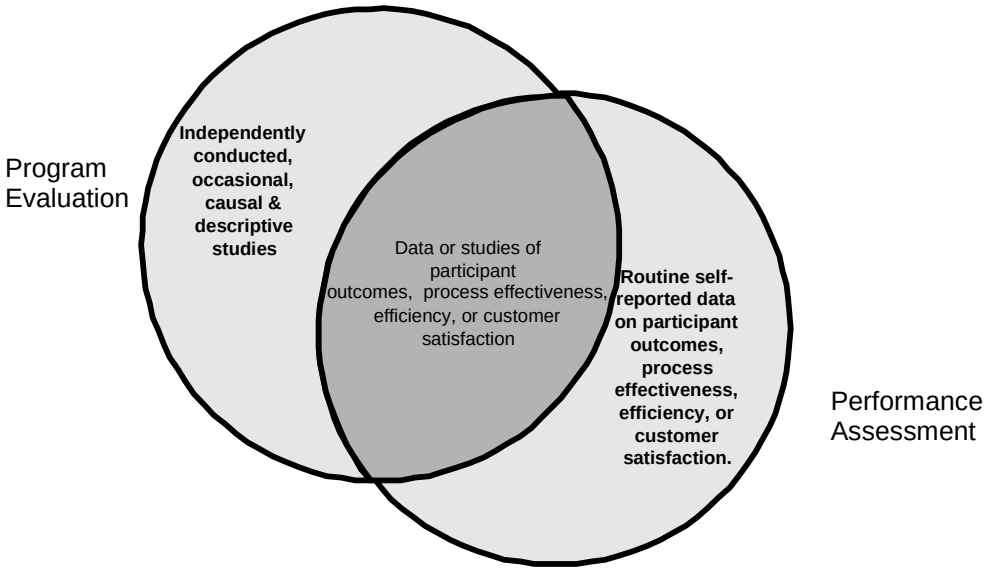
Now, to meet the increase in demand for program results, we try to maximize the use of our resources to identify and use information about program operations and program results so that we can focus on program measurement and program improvement. We are using existing information systems at the state and local levels to yield data on program results, and we have begun to develop partnerships between our evaluation office and our program offices to integrate the varied forms of performance information for decision makers.

Evaluations are systematic analytical efforts that are planned and conducted in response to specific management questions about performance of programs or activities. Unlike performance monitoring, which is ongoing, evaluations are intermittent and conducted when needed. Evaluations often focus on why results are or are not being achieved, or they may address issues such as relevance, effectiveness, efficiency, impact, or sustainability. Often, evaluations provide management with lessons and recommendations for adjustments in program strategies or activities.

Performance monitoring systems track and alert management whether actual results are being achieved as planned. They are built around a hierarchy of objectives linking activities and resources to intermediate results and strategic objectives. For each objective, one or more indicators are selected to measure performance against explicit targets (planned results to be achieved by specific dates). Performance monitoring is an ongoing, routine effort requiring data gathering, analysis, and reporting on results at periodic intervals.

A pictorial representation of the relationship between evaluation and performance measurement, in which each activity has its unique characteristics but they overlap to produce a complete range of information needed by program managers, appears as Figure 1.

Figure 1



While performance monitoring and evaluation are distinct functions, they can be highly complementary if they are appropriately coordinated with each other. Evaluations should be closely linked or integrated with performance monitoring systems. Performance monitoring will often trigger or flag the need for an evaluation especially when there are unexpected gaps between actual and planned results that need explanation.

ED needs to know not only what results were achieved (via the assessment system) but also how and why they were achieved, and what actions to take to improve performance further (via evaluation). Thus, evaluation makes unique contributions to explaining performance and understanding what can be done to make further improvements. Evaluation is an important, complementary tool for improving program management.

Evaluations serve five major roles in the GPRA environment:

1. **Evaluations provide information beyond performance measures.** Data obtained for reporting progress or performance may leave information gaps that evaluations can answer. Questions of causal relationships and certain types of programmatic effects cannot be answered with annual performance data.

2. **Evaluations validate performance data and refine performance indicators**. Performance data are drawn from many sources and do not exhibit the same degree of statistical reliability that evaluations offer. An evaluation can validate, or serve as a benchmark, for performance data as well as serving to refine indicators.

3. **Evaluations address strategic, not programmatic, goals**. Performance measures can address program goals, but they cannot always address strategic goals. Evaluations can focus on strategic goals and, sometimes by incorporating performance data from several programs that address a common goal, they can offer more complex inferences and understanding by being able to address a breadth of programmatic experience. Program evaluations also serve as valuable supplements to program performance reporting by addressing policy questions that extend

beyond or across program borders, such as the comparative advantage of one policy alternative over another.

4. **Evaluations guide program improvement**. Performance data provide useful and valuable information to program managers to improve program administration. Evaluations, frequently taking a broader and/or more in-depth approach to program structure and results, yield information that can lead to program improvement strategies and might address statutory, regulatory, or administrative changes based on methodologically sound grounds.

5. **Evaluations are long term**. Evaluations can explore hypothesis and present information beyond the experience of individual education programs. Some evaluations are designed to present information to aid Congress in the reauthorization of education laws. Evaluations are frequently used to advise departmental management and have implications for budget decisions. Evaluations are not as focused on real-time data as performance measures are, but there are opportunities for increasing congruence between evaluation and performance measurement.

Evaluation Directions for ED in the GPRA Environment

To fulfill our requirements to evaluate the effectiveness, quality of implementation of programs, and program results, we are continually analyzing our evaluation procedures to provide policy-relevant information in an effective, efficient, and cost-effective manner. In last year's Performance Report to the Congress, we reported on the revamped evaluation strategy that ED was undertaking. The reinvention of our evaluation processes continues, and we have refined our evaluation goals and principles for the coming year.

The following evaluation principles describe our evaluation goals that will produce credible and policy-relevant information for educational decision makers and the Congress:

Support Performance Measurement

- **Apply the GPRA requirements to reinforce development of performance measures to assess program outcomes and implementation quality on a regular basis.** GPRA explicitly reinforces our use of program evaluations to obtain objective measures of program results. Evaluations serve to check on program or other performance data and to provide causal explanations for observed performance not obtainable through performance measurement systems.

- **Continue to use multiple measures to assess and validate the consistency of evaluation results.** Confidence in evaluation results is greatly enhanced when corroborated across multiple studies rather than a single study. ED's large Title I study was particularly effective in applying a multiple-study design to assess and corroborate student outcomes from different information sources, including National assessments, individual state assessments, and urban district assessment results.

- **Develop and use performance benchmarks as a way to provide common evaluation metrics across diverse state and local systems.** In some areas, our studies have provided rich, generalizeable information, such as in the examination of the targeting and use of Federal resources. However, our studies sometimes lacked performance benchmarks against which to judge the quality of implementation in program activities. An explicit set of performance benchmarks is sometimes needed to judge the quality of program practices and results. When

evaluations have these quality benchmarks, they focus information collection. When benchmarks do not exist, evaluations need to launch developmental work to specify them.

Improve Measurement and Methodology

- **Collect rigorous, evidence-based data rather than relying on self-reports.** Many of our surveys of education professionals provided descriptive information on numerous important questions, such as hours of professional development or numbers and roles of teacher aides. While relying on school staff judgments about implementation of content and performance standards or the alignment of instruction with assessments is valuable, the responses are likely to lack the objectivity necessary for data reporting. For example, while we may have trend data on principals familiar with and indicating alignment of their instruction with standards, socially desirable responses may always result in inflated percentages reporting familiarity and alignment. Evidenced-based responses that reflect in-depth observational information and the use of more sophisticated questionnaires for obtaining factually based information are required. For example, teacher-time-use estimates have been shown to be reasonably reliable measures of teachers' actual time use.

- **Make greater use of causal methodologies, especially to evaluate instructional practices.** The primary causal evaluation model we have supported is the large-scale longitudinal study of schools. With independently administered assessments and in-depth information on the effects of program interventions, these studies have substantial potential to provide information on what works in the school or classroom. Other school-level information collected in the past was, for the most part, descriptive information of current practices. Descriptive information is valuable, but it is not sufficient to add to the knowledge base about the effectiveness of particular instructional or other practices for at-risk populations. Future evaluations need to place more emphasis on causal evaluations of an experimental or longitudinal nature of specific interventions.

Use Technology to Improve our Response Time

- **Take advantage of the availability of information from other evaluations of systemic reform and general-purpose data sources, in addition to ED evaluations.** States produce regular student assessments that provide massive amounts of information to evaluate Title I and other Federal programs. Statistical agencies collect general benchmarks against which to measure Title I outcomes and implementation. Foundations support systemic reform and educational innovation activities. Research on systemic reform and related interventions can reinforce evaluations on effective practices. Evaluations should develop information banks and other knowledge-management strategies for these different information sources.

- **Collect data electronically to provide real-time information.** New electronic methods provide opportunities to speed up data collection and increase accuracy. States already have considerable information on Web sites that could be harvested far earlier than when formally collected through state performance reports. ED is piloting with two states an Integrated Performance Based System to electronically harvest state and local education data for Federal analysis and use.

- **Develop a management information system to integrate evaluations, program monitoring data, and general-purpose data collections from across the Department.** Many of our data collections operate independently and fail to build on one another. We are developing a mechanism for integrating information from these multiple sources, which will strengthen the abilities of program offices to provide technical assistance to states and districts.

Improve Capacity Building

- **Reinforce training, capacity building, and the introduction of systemic evaluation and assessment procedures throughout ED to make GPRA an essential component for all program plans**. Not all ED staff have the awareness that GPRA is a long-term commitment that affects everything we do. Not all offices have grasped the implications of having concrete performance goals and targets, being accountable for these goals and targets, and reporting annually. Programs face a steep learning curve and knowledge gaps; our training activities for fiscal year 2000 and beyond will meet this need.

- **Use Federal evaluations to feed information back, and provide evaluation tools to improve evaluation capacity and use at state, local and national levels.** Evaluations can become more powerful change agents if they can build the capacity of different levels of program operations—state, local, school, and Federal—to provide the information each level needs to continuously reflect and improve results. Building capacity for feedback and reflection would strongly reinforce continuous improvement provisions underlying ED programs.

These principles reflect ED's continued focus on improving evaluation strategies, undertaken in concert with GPRA, to obtain and make credible, reliable, and timely information available to decision makers.