MEASURING THE PERFORMANCE OF INTERNATIONAL EXCHANGES & TRAINING PROGRAMS

INTERAGENCY WORKING GROUP ON U.S. GOVERNMENT-SPONSORED INTERNATIONAL EXCHANGES AND TRAINING (IAWG)

AUGUST 2000

Overview

The term "performance measurement" has been echoing through the halls of federal organizations for several years. Organizations strive to incorporate this and other principles of results-based management into their corporate cultures and day-to-day activities. Yet understanding, developing, and implementing a basic performance measurement system continues to challenge many program administrators.

Executive Order 13055 and subsequent legislation require the Interagency Working Group on U.S. Government-Sponsored International Exchanges and Training (IAWG) to develop recommendations on performance measures for all such exchanges and training.[1] The IAWG views this mandate as an opportunity to explore and understand performance measurement and develop approaches that will benefit the international exchanges and training community. While the various agencies that implement international exchanges and training programs have adopted performance measurement to differing degrees and with various levels of success, there is a rich variety of experiences that can and should be shared among agencies.

Often, federal government performance measurement is linked to the Government Performance and Results Act of 1993 (referred to commonly both as GPRA and the Results Act), which requires, among other things, every federal agency to submit annual performance plans to Congress. But the principles of performance measurement and results-based management go far beyond this monumental piece of legislation. Performance measurement is a hallmark of sound management, at all levels. It provides managers with the tools to demonstrate what works and why, and to make direct linkages between resource allocations and results. Performance measurement, when done correctly, contributes to a systematic approach to achieve program goals and document success.

Written material on "managing for results" abounds. This body of work, however, gets mixed reviews from the user community with regard to clarity and usefulness. Many useful performance measurement case studies exist in available literature, but few focus on international exchanges and training programs. While several "primers" discuss lessons learned and best practices for developing performance measurement systems, none specifically address the challenges of international exchanges and training programs, and many contain seemingly conflicting advice.

In an effort to alleviate this shortage, assist its member organizations, and begin to meet its congressional and Executive Branch mandates, the IAWG presents this performance measurement report, broken into three sections:

- Section I acts as a primer on performance measurement. After reviewing and synthesizing various resources on performance measurement, the IAWG developed guidelines specifically geared to international exchanges and training programs. Two

[1] While the IAWG addressed performance measurement in its first two Annual Reports, this is the first full report on performance measurement, as required by legislation. The IAWG is not going to address the creation of full performance measurement systems, but will provide general information to assist in that process.

separate surveys of organizations implementing international exchanges and training systems show that few measure performance in a meaningful way. Performance measurement is still a relatively new and unpracticed concept in the international exchanges and training arena. The majority of this report focuses on this review, because a solid understanding of performance measurement is necessary before any further work in the area can occur.

- Section II presents two profiles of IAWG member organizations that have taken innovative steps to implement performance measurement systems. The IAWG hopes that these profiles will provide ideas and inspiration to those organizations facing performance measurement challenges.

- Section III discusses the feasibility of developing common performance measures among international exchanges and training programs and suggests approaches to grouping programs to facilitate cross-program measurement.

This report draws material and inspiration from various resources on performance measurement. For additional information on these resources, see Appendix 4.

Section I. Performance Measurement Primer

Why Measure Performance?

"Because the Results Act tells us to" is a frequent response to this question. But there are many other reasons to measure performance that exist outside the mandates of the Results Act, and will continue to exist regardless of its status. Performance-based management incorporates some of the most basic tenets of good management; it provides accountability, facilitates teamwork, and helps motivate employees.

Performance measurement gives managers the tools they need to achieve programmatic and organizational goals. Data gleaned through the performance measurement process provide more objective than anecdotal information, and are less subject to variations in interpretation. Such data enable managers to justify their decisions to stakeholders.

When implemented properly, performance measurement can be used to improve communication among program stakeholders and to develop and strengthen partnerships among programs and organizations with similar goals and objectives. Performance measurement also provides tangible feedback to employees. Results demonstrate how various approaches and delivery mechanisms contribute to overarching organizational goals.

Understanding the Terms

Agencies must first understand the terminology associated with performance measurement before attempting to adopt it. A definite results "lingo" permeates existing resources. Appendix 3 of this report contains the "official" terminology from OMB Circular A-11, which is the Office of Management and Budget's interpretation of and guidance for compliance with the Results Act. However, some variations exist between the A-11 terminology and that found in other performance measurement resources. Although Circular A-11 serves as the primary reference for government users, it continues to evolve, as evidenced by a recent revision.

Agencies must understand the following terms before devising or implementing effective performance measurement systems

Mission: The reason for an organization's or program's existence. A few resources also refer to mission/objective statements in a more general sense. The United States General Accounting Office (GAO) notes that the mission statement "brings an Agency into focus…explains why the Agency exists, tells what it does, and describes how it does it."[2] The term "**mandate**" may also be used to describe specific directives contained in a mission statement or other mission-oriented documentation.

Goals: The desired end states toward which activities are directed. Goals determine how an organization will execute its mission over time. Although some entities use "goal" and "**objective**" interchangeably, an "objective" usually describes a more specific level of achievement than a goal. Specific statements of objectives identify desired outcomes. (This definition of "goal" corresponds with the term "**general goal**" in OMB Circular A-11.)

There are four basic types of performance data (inputs, processes, outputs, and outcomes) that are referred to when planning and conducting performance measurement activities:

Inputs: The resources expended on a given activity or program. Funding, human resources, and equipment constitute the three most common inputs. Budgeted and allocated inputs are not linked to actual results, so only expenditures should be included as inputs in results reports. When developing out-year performance projections, however, expenditure data are not yet available, so budget amounts may be used.

Example 1

To provide an example of the terms reviewed in this section, the IAWG performance measurement study group created an example for a Federal Aviation Administration (FAA) program designed to train foreign air traffic controllers to use new and advanced control equipment:

Mission: "…To train foreign nationals…in aeronautical and related subjects essential to the orderly and safe operations of civil aircraft."

Program Goal: To improve safety through use of more accurate and reliable control equipment internationally.

Inputs: The funding for the program and trainers' time.

Process: Contract an external organization to perform training, and use a combination of formal classroom and on-the-job training.

Outputs: The number of controllers trained and the number of training sessions held.

[2] General Accounting Office (GAO), *Executive Guide: Effectively Implementing the Government Performance and Results Act*, GAO/GGD-96-118, 1996, p. 13.

Processes: Delivery mechanisms and amounts of activity undertaken to carry out programs and activities. Process evaluations contribute to internal management decision making.

Outputs: The products and services produced in implementing a program. Most federal entities, indeed even the IAWG, report output information such as the number of participants on a given exchange program, the number of trainees per session, or the number of training seminars offered.

Outcomes: Results often expressed as conditions, behaviors, and attitudes.[3] Outcomes further break down into two categories: **intermediate outcomes** (the intermediate results along the way to achieving final desired results/outcomes) and **end outcomes** (the end results of a program activity vis-à-vis its goals).

For federal performance measurement, GAO and OMB stress outcomes over other data. However, performance measurement literature (outside the scope of the federal government) promotes more of a balance among inputs, processes, outputs, and outcomes, when looking at what to measure. Many program managers measuring performance often confuse outputs and outcomes. Good program management requires monitoring and measuring outputs; with few exceptions, however, these pieces of information alone do not constitute a program's results.[4] Although key components to performance measurement include monitoring and recording output levels, performance reporting should include all categories of data.

Example 1 (cont.)

Intermediate Outcome: The trainees' satisfaction with the training and the knowledge and ability to use the new equipment.

Performance Measure: Percentage of trainees expressing a high level of satisfaction with the training (score of 4 or 5 on a five-point scale) and the percentage of trainees that pass an equipment usage proficiency test.

Target: Seventy-five percent of trainees rating satisfaction level of 4 or 5 and 80 percent passing proficiency test.

End Outcome: Decreased aircraft accidents/incidents blamed on controller error.

Performance Measure: The percentage decrease of aircraft accidents/incidents blamed on controller error over a one-year period following the training program.

Target: A 10 percent decrease in the percentage of aircraft accidents/incidents blamed on controller error over a one year period following the training program as compared to the average number of the same types of incidents from the previous three years.

The example above is hypothetical, though the Mission is that of the FAA's international exchanges and training programs.

[3] Harry Hatry further illustrates by stating, "outcomes are the events, occurrences, or changes in conditions, behaviors, or attitudes that indicate progress toward achievement of the mission and objectives of the program." *Performance Measurement: Getting Results*, The Urban Institute Press, Washington, D.C., 1999, p. 15.

[4] Hatry notes that the category of the performance information may depend on the perspective of the agency (p.12). For instance, an overseas educational advising center develops an informational video for use by prospective foreign students. This video may be deemed an output for the contractor, but an intermediate outcome for the funding federal agency.

Performance Measure: A particular value or characteristic used to measure output or outcome. This term is used interchangeably with the term "**performance indicator**."

Target: Specific, planned level of result desired within an articulated measure/indicator. (OMB Circular A-11 refers to targets as "**performance goals**" and further breaks them down into two categories: "**output goals**" and "**outcome goals**.")

Each of these terms will be discussed in more detail below, using exchanges and training program-specific examples.

Define the Mission

Good performance measurement begins with a clearly-defined mission. A program mission explains **what** (the activity) is done for **whom** (the customer/beneficiary) and **why** (the purpose/goal). Missions can be culled from a wide variety of sources, including but not limited to:

- Legislation, executive documentation (Executive Orders, Presidential Memoranda, etc.), and regulations.
- Organizational planning documents (budget documents, strategic plans, etc.).
- Interagency agreements and memoranda of understanding.
- International agreements.
- Program descriptions and annual reports.
- Consultations with stakeholders.
- Other government organizations with similar programs.
- Program evaluations and audits.

Organizations should consider several points when developing mission statements:

- Mission statements should be inspirational and encompass the major, overarching result(s) sought.
- Missions are not necessarily static and can change based on foreign or domestic priorities. (However, stable missions and goals facilitate the creation of trend data.)
- Mission statements should not include numerical targets, which change with time and programming phase.

Each organization's performance measurement system starts with strategic plans, as defined and required by the Results Act. While most sources recommend that the missions of all programs relate to the strategic plan, this is not always possible with international exchanges and training programs. An organization's mission statement constitutes the foundation of its strategic plan. However, references to specific international program missions may be excluded from the agency-level strategic plan, especially for non-foreign affairs agencies and for those agencies that conduct international exchanges and training programs using funds transferred from another federal agency. The international programs within these agencies may represent too small a proportion of their activities to be represented.

Outline Goals and Objectives

Goals and objectives address the mission statement/mandate, articulate desired results, and form the basis for every subsequent element of the performance measurement system. They may be related to process, outputs, and/or outcomes. Goals and objectives should be measurable and within the realm of control of the organization establishing them. They also need a strategic rationale (i.e., the goals and objectives should be logical and appropriate). Both the mission statement/mandate and any relevant legislation or interagency agreements should be reviewed to determine if goals and objectives are explicitly stated.

Common goals and objectives among programs can facilitate the application of similar performance measures. When measuring and reporting performance, organizations can group these programs together or sample data from them to demonstrate how multiple programs, targeting varying audiences and/or using different delivery mechanisms, can work together to fulfill a single or similar goals. By extending this principle even further to include programs from other government and nongovernmental organizations, managers can continuously monitor for duplicative programming and demonstrate inter-organizational program complementarity. (See page 14 for an explanation of performance partnerships.) As Congress demonstrates increased interest in programmatic duplication, it is in an organization's best interest to enhance communication and coordination among programs with similar goals and objectives.

As an example, international exchanges and training programs funded under the Freedom Support Act (FSA) lend themselves to a certain degree of grouping within or among certain organizations. These FSA programs all share one common goal: to promote the transition to democracy and market economies in the New Independent States (NIS) of the former Soviet Union. FSA-funded programs, which exist throughout the U.S. Government, reach a wide variety of audiences (from high school students to mid-career professionals to senior researchers) and employ a multitude of methodologies (study abroad, in-country training seminars, observation tours, etc.). Regardless of their distinctions, these programs all have the same overarching goal. (It should be noted that the fact that they share a common funding source facilitates the grouping of these programs.) Organizations can group these programs together for some end outcome measurement and can benchmark against other federal organizations with similar FSA-funded programs. To economize resources, end or intermediate outcomes could be sampled from related programs.

The terms goal and objective can be used interchangeably or, alternately, reflect a nuanced level of specificity. Specific objective statements can be used interchangeably with "outcome statements," which are explained in the next section.

Define and Measure Outcomes

Organizations develop programs to fulfill their articulated missions, goals, and objectives, and specify desired outcomes for each of these programs. Outcome (or results) statements should be tightly focused to facilitate developing indicators. Otherwise, it will be difficult to develop indicators and maintain the focus of the performance measurement exercise. Also, broad outcome statements can lead to differences in interpretation.

The United States Agency for International Development (USAID) is one of the few foreign affairs agencies to systematize the performance measurement process and incorporate guidance and reference materials on the process into its body of organizational documentation. USAID's reference materials can benefit other government administrators of international exchanges and training programs. TIPS sheets developed by USAID's Center for Development Information and Evaluation contain institutional guidance. (These documents can be found at http://www.dec.org/usaid_eval/ under CDIE Performance Monitoring and Evaluation Tips.)

The following four USAID points on establishing outcomes and indicators are also useful in preparing outcome statements:[5]

- Be clear about the type of outcome your indicators imply (e.g., a change in situation, condition, the level of knowledge, attitudes, or behaviors of the target group).

  To illustrate, USAID uses democratization programs as an example: "...changing a country's law about voting is very different from changing citizens' awareness of their right to vote, which again is different from their voting behavior. Each type of change is measured by different types of indicators."

- Indicate the level of change desired:

  Absolute change -- the creation or introduction of something new

  Relative change -- increases, decreases, improvements, strengthening or weakening in a currently existing state, but at a higher or lower level than is considered optimum

  No change -- the maintenance, protection or preservation of something that does not warrant change

- Indicate the target group: individuals, groups, communities, organizations, systems, geographic regions, industries, etc. (USAID refers to this as identifying the "unit of analysis.")

- Define the expected relationship between activities and results. (This is a critical point raised throughout different performance measurement resources in various contexts. See below for further explanation.)

Hatry suggests that individuals developing outcome statements also ask the following questions:

- Do outcomes address all the objectives identified in the mission and goals statements?

[5] U.S. Agency for International Development (USAID), Center for Development Information and Evaluation, *Selecting Performance Indicators*, Performance Monitoring and Evaluation TIPS, No.6, 1996.

- What would be the effect on customers/stakeholders if the program's budget and resources were substantially increased, decreased, or eliminated?

- Are there any [unintended negative] consequences or effects associated with the program that should also be monitored? (If these can be tracked on a regular basis, they should be included as [unintended] outcomes.)[6]

Program staff should differentiate between intermediate and end outcomes. They need to base the classification of outcomes on the reality of the results environment. Other factors often affect end outcomes, which makes them more difficult to measure. End outcomes often do not occur until long periods of time have passed, especially in the case of many exchange programs. For instance, the "mutual understanding" programs of the State Department or the basic research programs of the various science agencies may take years to demonstrate end outcomes, which may not result solely from programmatic interventions. Intermediate outcomes can often be achieved and measured in the short-term. They also are less likely to be affected by external factors. Program delivery mechanisms, however, often affect intermediate outcomes.[7]

Measurement approaches should be adapted to fit the type and level of outcome anticipated. Organizations may wish to consider creating a measurement pyramid. At the broadest level, outputs for all programs and activities are measured and recorded at the conclusion of each program and activity. This provides basic management data. At the mid-level, programs are grouped by either delivery mechanism or general objectives. For instance, all programs using in-country, train-the-trainer delivery mechanisms may have similar desired intermediate outcomes, such as participant satisfaction with training offered and development of training materials to be used in subsequent training sessions. These intermediate outcomes may not need to be measured for every program using this approach, so a somewhat smaller sampling can be used to report outcomes annually. At the narrowest level, programs with similar desired end outcomes could be grouped together with an even smaller number sampled at a later date (1-2 years following the conclusion of the program). For end outcomes, regular measurement at early intervals following a program provides little value.

To understand the relationship of a program to its desired outcome, it is useful to understand how uncontrollable factors may affect the outcome. Program staff should specifically identify, and rigorously define, as many external factors as possible. Intermediate indicators can be developed to monitor these elements as well, even though they cannot be influenced. These indicators can be presented along with program performance indicators to provide context for how positively or negatively external factors interact with the program's aim.[8]

---

[6] Hatry, p. 53.
[7] In relationships with contractors and grantees, the degree of performance measurement that can be accomplished may depend in part on the duration of the contract. End outcomes (and even many intermediate outcomes) may not be apparent during the period of the contact or grant.
[8] The 1993 GPRA legislation states that strategic plans should include, "an identification of those key factors external to the agency and beyond its control that could significantly affect the achievement of the general goals and objectives…," Government Performance and Results Act, PL 103-62, Section 3, para. (a)(5) or U.S.C. Title V, Chapter 3, Section 306, para. (a)(5).

The farther in the future an outcome occurs, the more difficult it is to determine the extent to which program activities caused the outcome and the degree to which external factors affect results. This challenge is more pronounced with international programs because program managers are often far removed from the environment in which outcomes will be realized and do not have the resources to fully evaluate causal relations. The degree to which program activities affect outcomes, or at least acknowledgement that the cause-and-effect relationship is unclear, should be included in program reports.

Proxy or surrogate performance data may be needed to measure outcomes that cannot be measured directly. For example, if a U.S. science organization provides fellowships to international experts to work on joint projects in the United States, it may be impossible to quantify the end outcome(s) of the research during the period covered by the grant. Surrogate information in this case could include peer reviews and the number of citations of project publications in technical literature.

Develop Indicators

Indicators (also called measures) tell program managers whether goals are being met. They should not be confused with targets, described below. Conceivably, there can be several indicators for every desired outcome. Indicators can be expressed in both quantitative and qualitative terms. Most organizations, though, will want to emphasize quantitative indicators where possible because they are less subjective and are more easily compared among programs and against baselines.

One common mistake program managers make is selecting indicators that are the easiest to measure. While feasibility and cost-effectiveness are important concerns, managers should develop indicators that represent the greatest value in determining program results and facilitating good program management. While it may not be possible to utilize all these indicators at the present, some may be useful in the future if resources change and/or data collection systems are adapted or improved. Additionally, “extra” indicators can be used for more in-depth program evaluations.

Example 2

Though challenging, there are ways to quantify seemingly qualitative information. The following, taken from the USAID TIPS sheet “Establishing Performance Targets," provides a useful example:

To measure an intermediate result that emphasizes improvements in quality of maternal and child health services, USAID/Yemen devised a scale that transforms qualitative information about services into a rating system against which targets can be set:

0 points = Service not offered
1 point = Offers routine antenatal care
1 point = Offers recognition and appropriate management of high risk pregnancies
1 point = Offers routine deliveries
1 point = Offers appropriate management of complicated deliveries
1 point = Offers post partum care
1 point = Offers neonatal care

$$\text{Score: } \frac{\text{Total actual service delivery points}}{\text{Total possible service delivery points}}$$

USAID, Center for Development Information and Evaluation, "Establishing Performance Targets," Performance Monitoring and Evaluation TIPS, No.8, 1996, p.2.

Indicators may change over time, based on the data collection and management resources available to the organization. Static indicators are not a prerequisite of a good performance management system. However, organizations must realize that trend analysis requires common indicators over a period of time. For organizations that are just developing performance measurement systems and that plan on augmenting their data collection and analysis capabilities in the future, a mixed approach is recommended -- use available data now and continue to measure against it in the future, even as you add more "sophisticated" indicators to the mix.

Indicators should be:

- Useful to managers in making program and budgetary adjustments (in both the short- and long-term).
- Directly related to the outcomes desired and the goals set.
- As removed from external influence as possible.
- As specific as possible.
- Organizationally feasible and within acceptable cost parameters.
- As objective as possible.
- Using reliable data.
- Clear as to the timeframe covered.
- Included if they are important to the program, even if they are not exhibiting change or show steady positive results.
- Unidimensional.

Most sources agree that managers should choose as many indicators as will be helpful, while being mindful of the quality of the data that can be collected and the organizational cost to collect data. The SMART acronym may be helpful in remembering guidelines for indicators: Specific, Measurable, Achievable, Realistic, Time-bound.

Example 3

Sample intermediate outcome indicators can be developed for several types of exchanges and training programs. Please note that the IAWG is not advocating the appropriateness of these indicators for all programs, but simply supplying samples that may be appropriate for different models:

Training Programs (U.S.-based and foreign):

- The level of satisfaction of trainees with training sessions or components (measured at the end of or immediately following the training session)
- Trainees' self assessment of the degree of impact training programs have had on their professional activities and/or competencies (measured after a period of several months to a year)
- Changes in scores on technical competency examinations (measured at the end of, or immediately following, the training sessions)
- The number of trainees who obtain special certifications (or degrees in the case of academic programs) at the conclusion of the program

Train-the-Trainer Programs (the above would apply along with):

- The lifespan of training materials generated through a program (measured in the mid-term following the program)
- The number of training sessions conducted or individuals subsequently trained by trainers trained under the auspices of the program within a specified time period

Research Programs:

- The number of citations in technical journals or comparable professional publications
- Assessments by peer review panels
- The degree to which findings are being disseminated
- The degree to which findings contribute to other research programs or product development

Exchange Programs:

- Exchangees' self assessments of degree of attitudinal change about targeted issues
- The frequency of communication between exchangees and contacts made during the exchange.
- The number of new projects initiated among exchangees as a result of the exchange

Establish Performance Targets

The Results Act requires that each organization's annual performance plan include annual performance goals (targets) and indicators for the fiscal year.[9] To keep from confusing annual performance goals with the general goals defined above, this report will address these types of narrow annual goals as "targets." While it is not advisable to include specific quantifiable targets in mission statements, they can be included in tandem with performance indicators. However, performance targets should not be confused with indicators. In its performance target TIPS sheet, USAID explains that while an indicator "defines how performance will be measured along

[9] Office of Management and Budget (OMB), *Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports*, Circular A-11, Part 2, 1999, pp. 485-486.

a scale or dimensions, the target identifies the specific, planned level of result to be achieved within an explicit timeframe."[10]

Hatry presents a useful series of suggestions to keep in mind when establishing targets:

- A target does not have to be a single value, e.g., it can be a range.
- Consider previous performance.
- Consider benchmarking against the best.
- If benchmarking against the best is considered too great a challenge, use the average performance of all units.
- Consider the outcomes achieved in the past for different customer workload categories.
- Consider the performance levels achieved by other jurisdictions or by private firms with similar activities and workload or customer compositions.
- Make sure the targets chosen are feasible, given the program's budget and staffing plan for the year.
- Identify any new developments -- internal and external -- that may affect the program's ability to achieve desired outcomes.
- Target setting for periods shorter than a year needs to be done in the context of seasonal factors.[11]

Data Collection, Validation, and Verification

Any given performance measurement system is only as strong as the data around which it is built. As has been noted previously, performance data can be either quantitative or qualitative, though a general preference for quantitative data is expressed in most sources. Quantitative data are seen as more "factual," are less vulnerable to subjective interpretation, facilitate aggregation and sampling, and lend themselves to cross-program comparison and analysis. Qualitative data, while acceptable to measure performance, are often best used to illustrate and provide context and depth of understanding to results.

A wide variety of sources for quantitative data exist (i.e., program reports, surveys/ questionnaires, focus groups, examinations, trained observer ratings,[12] etc.). Regardless of the source used, organizations must have a system in place to record the data, organize it for analysis, and archive it for trend analysis. This has presented a challenge to many. Very few federal agencies have reliable data management systems in place.[13] Implementing a sound performance measurement system requires adequate data management systems.

The annual performance plans developed by agencies must identify means by which agencies will verify and validate performance data. GAO defines verification as the "assessment of data completeness, accuracy, consistency, timeliness, and related quality control practices" and

[10]USAID, Center for Development Information and Evaluation, *Establishing Performance Targets*, Performance Monitoring and Evaluation TIPS, No.8, 1996, pp. 1-3.
[11]Hatry, pp. 129-130.
[12]Hatry goes into some detail on use of trained observer ratings on pp. 86-93.
[13]See IAWG FY 1998 Annual Report, Chapter 2, Section 2 , "Data Management," pp. 21-31.

validation as the "assessment of whether data are appropriate for the performance measure."[14] Adequate data validation and verification enhance the credibility and value of agency performance data and facilitate the communication of this data to stakeholders.

Develop Reporting Strategy

Performance reports range from simple internal reports between program managers to the formal annual performance reports to Congress. There are as many potential target audiences as there are circumstances under which reports may be generated. Regardless of the reason for the report and the intended audience, several recommendations mentioned in performance measurement sources can assist federal managers:

- **Clarity and Conciseness:** Make performance measurement reports clear and concise. Almost all sources suggest that drafters of performance reports focus on the intended audience, select a few indicators on which to report, eliminate extraneous detail from the report itself or move it to an appendix, and be as concise as possible. Senior-level decision makers do not have the time or level of interest to wade through scores of indicators and all the collected data. They want to see the big picture as clearly as possible. For lengthy performance reports, include a one-page executive summary to make key information easily accessible.

- **Explanatory Information:** Include explanatory information in performance reports to reduce the potential for decision makers to misinterpret data. This is especially important if performance information reflects shortfalls and/or limited data sources. This information can be administrative (i.e., explaining resource limitations/changes, etc.) and/or address those external influences that are beyond the control of the program, yet can still affect outcomes.

- **Actions:** Strengthen reports by including a section on how the organization intends to use performance information. Address how resource allocation, delivery mechanisms, customer/stakeholder relations, and other factors will change based on performance data.

Additional Suggestions

Planning: To achieve an efficient, effective, and sustainable performance measurement system, organizations develop an internal performance measurement plan that systematizes the above and assigns responsibility for a) identifying and adjusting mission and goals, b) choosing and developing indicators, c) collecting data, d) analyzing data, e) reporting data, and f) implementing programming and budgetary adjustments in response to results.[15]

[14] GAO, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GAO/GGD-99-139, 1999, p.12.
[15] USAID's Center for Development Information and Evaluation's *Preparing a Performance Monitoring Plan*, Performance Monitoring and Evaluation TIPS, No.7, 1996, serves as an excellent resource in this area.

In creating performance measurement plans, agencies need to provide *clarity* to the performance measurement process, enhance *communication* among employees and managers responsible for various aspects of performance measurement, and provide organizational *continuity* to protect against staff turnover.

Stakeholders: Performance measurement does not occur in a vacuum. Many other factors can affect the viability and sustainability of international exchanges and training programs. Organizations must involve stakeholders (both internal and external) along the way to achieve broad consensus and buy-in. As federal managers know, the survival of international exchanges and training programs depend heavily on broad-based support for them. Stakeholders include, but are not limited to:

- Participants
- Program staff
- NGO and private sector partners
- Congress
- Executive branch organizations
- Overseas staff
- Foreign government representatives
- Taxpayers

Beyond stakeholder involvement is the notion of "**performance partnerships**." Performance measurement improves when organizations seek out other government and private/non-profit programs with similar goals and objectives to develop partnerships in the area of performance measurement. This specific type of partnership exists when organizations form close working relationships to discuss common goals and objectives, coordinate programming, share best practices with regard to approaches and methodologies, and develop similar or common performance indicators. This type of cooperation on program implementation and performance measurement can vastly improve end results.[16]

Performance partnerships could be especially beneficial in international exchanges and training programs where one government organization funds activities implemented by another organization. The IAWG's *FY 1998 Annual Report* noted that responsibility for performance measurement is one of the many challenges faced by organizations dealing with these "budget transfers." It is often unclear whether the transferring or receiving (implementing) agency is responsible for performance measurement and reporting. Each organization may have certain responsibilities for reporting outcome data, but may have dissimilar approaches to developing indicators, collecting data, and reporting outcomes. Three particular issues were raised in the IAWG's report:

[16] Hatry writes, "Partnerships are warranted when the program believes that desired outcomes would be best achieved by obtaining voluntary agreement among organizations as to a) the outcomes and outcome indicators to be tracked; b) how the data should be collected; c) the short- and long-term targets for each outcome indicator; and d) the roles and responsibilities of each organization in providing the particular service…Such agreements are called performance partnerships. They are a new concept and require significant time and effort to work out with other organizations." (p. 47)

- Over time, more and more budget transfers have been executed as Section 632(a) transfers.[17] Under this type of transfer, the transferring agency shifts virtually full accountability for the funds, including GPRA reporting, to the receiving agency. Transfer agreements under Section 632(a) generally require only quarterly or semi-annual reporting to the transferring agency.

- Foreign affairs agencies may transfer funds to agencies without foreign affairs mandates, but with areas of specific programmatic expertise. These recipient agencies have performance plans that reflect their domestic mandate and priorities. With programs conceptually originating in a foreign affairs agency and program implementation occurring in a domestically-mandated agency, there may not be clear agreement on project objectives in the foreign policy context.

- It is not a priority for the recipient agency to measure and report outcomes to Congress for programs for which it does not receive a direct appropriation and which may, in some instances, fall outside the parameters of its strategic goals.

Performance partnerships among these organizations would address these issues. Similarly, organizations that provide contracts or award grants for the implementation of international exchanges and training activities could improve their overall performance measurement system and ability to obtain desired results by establishing performance partnerships with their contractors and grantees.

Program staff should also be recognized as important stakeholders. Organizations often focus attention on decision makers and funders when communicating throughout the performance process. While these individuals are critical to an organization's programs, internal organization partners should never be overlooked. Performance information and reports should be shared with program staff on a regular basis. Internally, the success of any performance management system depends upon a team approach. Program staff should be given input at all levels and be given the opportunity to add explanatory information to reports, when warranted.

Examples, both real and hypothetical, of performance measures developed for international exchanges and training programs appear in Appendix 2.

[17] Foreign Assistance Act of 1961, PL 87-195, section 632 (a), p. 285. Also 22 U.S.C. 2392.

Summary

The major areas discussed above constitute traditional procedures for measuring performance.

1. Define mission (organization and/or program).

2. Outline goals that enable the program(s)/activity(ies) to achieve its/their mission(s).

3. Articulate program outcomes that will contribute to achieving goals.

4. Select indicators to determine if and to what degree outcomes (intermediate and end) are achieved.

5. Collect, validate, and verify data: Determine if data are appropriate, reliable, and credible and if they can be feasibly and cost-effectively collected.

6. Develop a reporting strategy.

7. Incorporate performance data and results into planning and budgeting processes.

When implementing the above:
- o Create a written plan/policy that articulates areas of responsibility.
- o Involve stakeholders and develop performance partnerships.

Section II. Performance Measurement Profiles

To reassess the performance measurement of U.S. Government-sponsored international exchanges and training programs, the IAWG Performance Measurement Study Group reviewed the results of the first performance measurement survey conducted by the IAWG in 1998, and sent a follow-up survey to those organizations that had responded that they measured performance to some degree. Sixteen follow-up surveys were distributed. Nine agencies returned survey responses, along with documents and other materials representing performance measurement approaches for 26 exchanges and training programs. The seven agencies that did not respond to the survey had either terminated their programs or acknowledged that they did not have a performance measurement system in place.

The IAWG Performance Measurement Study Group reviewed the material submitted in response to its follow-up performance measurement survey and selected two examples of organizations using innovative and pro-active approaches to adopting practical and effective performance measurement systems.

PROFILE I: USAID's Institutional Emphasis on Performance Measurement

The U.S. Agency for International Development (USAID) has taken great strides in reorganizing under a results-based management framework. USAID's reengineering is laudable in that its approach has been systemic, and the tools needed to accomplish the transition to results-based management have been widely available. The reengineering process directly affects USAID's international training activities, which support the achievement of all USAID strategic objectives.

In 1993, USAID became a reinvention laboratory under the National Performance Review (NPR) and volunteered as an early implementer of the Government Performance and Results Act. USAID leadership initiated broad management reforms to streamline operations and make them more efficient. The centerpiece of these management reforms is a results-driven strategic planning process.

Under this revised strategic planning process, USAID identified five strategic goals and 19 supporting objectives. Each USAID operating unit must target all of its activities to one or more of these goals and objectives. Annual milestones are established for each strategic objective and are used by USAID/Washington to assess progress and program performance.

By the end of 1995, virtually all operating units of USAID had strategic plans with explicit strategic objectives, performance indicators, and performance targets.[18]

USAID Strategic Goals

Goal 1: Broad-based economic growth achieved

Goal 2: Sustainable democracies built

Goal 3: World's population stabilized and human health protected in a sustainable fashion

Goal 4: Environment managed for long-term sustainability

Goal 5: Lives saved, suffering reduced, and development potential reinforced

Performance-based approaches appear in three major ways in the area of international training:

Guidance: The Performance Monitoring and Evaluation TIPS publications (13 issued in the last 5 years), produced by the Center for Development Information and Evaluation, provide general reengineering guidance. These guides give practical advice and examples to programmers and create and articulate a standardized approach throughout the agency. The IAWG used several of these TIPS sheets to prepare this report (see Appendix 4 for titles). In addition, USAID rewrote its operating handbooks to incorporate reengineering core values, chief among which is performance-based programming.

Best Practices: Training Best Practices modules incorporated the most successful practices from missions with the state-of-the-art thinking from training organizations and the U.S. private

[18]Background information on USAID's reengineering and strategic planning was taken from USAID's 1997 Congressional Presentation, which can be found at http://www.info.usaid.gov/pubs/cp97/main/cp-4.htm

sector. Although operating in greatly differing contexts (developed vs. developing countries), the same generic problems abound. USAID training planners have learned much from the more holistic approaches pioneered by the private sector.

Data Management: USAID's Training Results and Information Network (TraiNet)[19] , a distributed management information system, supports the planning and monitoring of agency-sponsored training of foreign nationals. TraiNet incorporates a results-oriented approach. Organized around a mission's strategic results framework, TraiNet encourages the user to approach training planning from the perspective of the sectoral results to which training contributes. TraiNet provides the strategic and reengineering framework on which training programs are designed and can be utilized by any training operation.

The TraiNet system makes great strides toward improving efficiency and communication in USAID. By implementing TraiNet, USAID eliminated at least five previously required forms, reduced data management duplication, and provided a systematic mechanism for cataloging results. TraiNet is also used to enter health insurance enrollment data for U.S.-based training participants and to prepare each Mission's Results Review and Resource Request (R4).

Challenges

Although USAID reinvented itself with relative success, its performance measurement system is still not fully implemented. Central training offices in most missions disappeared because of severe cutbacks in staff. Responsibility has devolved down to the strategic objective teams, in keeping with reengineering principles, but such teams consist of technical specialists who often have no background in using training to improve performance. Cutbacks make it more difficult to induce implementers to apply best practices that appear to be more labor-intensive.

Many other challenges exist to fully implementing performance-based training programs. Some USAID staff and host country partners still possess a "training is itself a good thing" mentality (also still seen in private sector managers). Some people view training as sowing many seeds in the hope that some will sprout. Also, real performance gap analysis can be a painful exercise. Planners may choose training when the real barriers to improved performance are not necessarily skill or knowledge gaps. Some use training to reward employees, who may have an agenda different from the goals of the program. Feeling empowered, well-trained and better-performing people may decide to jump ship, thereby leaving the organization in even worse shape.

PROFILE II: The Department of Education's Evaluation for Exchange, Language, International, and Area Studies (EELIAS) System

EELIAS is the electronic performance measurement system of the U.S. Department of Education's International Education and Graduate Programs Service (USED/IEGPS). USED/IEGPS administers 14 Title VI/Fulbright-Hays programs designed to enhance the

[19] A review of the TraiNet system was included in the IAWG's *FY 1998 Annual Report*, Chapter 2, Section 2: Data Management.

capability of U.S. higher education to provide the human resources, knowledge, and information necessary for national security, including economic well-being. The EELIAS system provides a data collection, management, and analysis tool to help USED/IEGPS ensure that its programs meet this critical mission.

IEGPS awarded an atypical five-year grant to the Johns Hopkins University National Foreign Language Center (NFLC) to develop EELIAS. When complete, EELIAS will include:

- Rigorous definitions for the overall Title VI/Fulbright-Hays mission and goals.
- Clear objectives (outcomes desired) that support each individual goal of all Title VI programs.
- A complete set of indicators and measurements related to each objective of every Title VI program.
- A precise mapping of how the 14 Title VI/Fulbright-Hays programs individually contribute to the overall mission..
- On-line, web-based data collection forms tailored to each Title VI/Fulbright-Hays program.[20]
- A web-based data warehousing and reporting system, fully compatible with existing USED information technology systems that contain other data EELIAS needs for measurement.
- Built-in analyses and reporting mechanisms to comply with annual GPRA requirements.

Title VI / Fulbright-Hays Mission:

To enhance the capability of U.S. higher education to provide the human resources, knowledge, and information necessary for national security, including economic well-being.

EELIAS-Specified Strategic Goals *(derived from the mission)*:

1. The generation of knowledge and information about the world, its languages and peoples
2. The development of experts on world languages, world areas, and international affairs
3. The training of business and other professionals in political, social and economic domains, capable of practicing their profession in any part of the world and interacting with representatives of any culture
4. The education of the citizenry, in order that they become cognizant of the global dimensions of national well-being and security
5. The warehousing, within the U.S. higher education, of a permanent capacity for the production and maintenance of the human resources and knowledge relevant to all areas of the world, including access to such knowledge, as is necessary for the national security

All the 14 IEGPS programs are being phased in to the EELIAS system over the five-year lifetime of the NFLC grant. Currently in its third year, the EELIAS system already includes many of these programs. The following steps are taken to phase-in programs on a rolling basis:

- First year: Program task force develops indicators and program-specific data reporting packages/instruments and implements on-line data collection.
- Second year: Collected data are reviewed, evaluated, and reported.
- Succeeding years: Data are reviewed, evaluated, reported, and responsibility is permanently shifted from the EELIAS project staff to IEGPS staff.
- NFLC will produce a final report of how well the 14 Title VI programs are meeting the national needs, after which IEGPS will have full ownership to continue EELIAS.

[20] The IEGPS and NFLC are currently awaiting OMB approval for the first EELIAS electronic data collection forms.

When complete, EELIAS will replace all current Title VI reporting requirements.

Developing EELIAS

NFLC formed task forces comprised of experts and stakeholders to guide the development process. Each task force included representatives from USED-IEGPS program management, the Office of Management and Budget (OMB), scholars from the various program fields (language, area studies, international education, and international business education), and nationally recognized experts in program evaluation. These members represented three primary constituencies: program administration (IEGPS), external oversight (OMB), and the Title VI grantee community (subject matter scholars). The program evaluation experts served to mold a performance system from this assembled expertise.

> **EELIAS Performance Indicator Development**
>
> The following is an example of a single set of EELIAS performance indicators, derived for a single program, the Foreign Language Area Studies (FLAS) program. The FLAS program addresses strategic goals 2, 3, and 5.
>
> **Performance Objective 1:** Broaden the range of disciplines and languages for which FLAS fellowships are awarded. (*This objective supports strategic goals 2 and 3 only. Other FLAS objectives include strategic goal 5)*
>
> **Performance Indicators** *(supporting Performance Objective 1):*
>
> *Indicator 1a:* Number of applicants for FLAS fellowships by language, by level of instruction, and by discipline
>
> *Indicator 1b:* Number of disciplines, levels of instruction, and languages in which FLAS fellowships are awarded

The task force examined each of the 14 programs, linked them to strategic goals, identified specific program objectives, and developed precise performance measures/ indicators (see example to the right). Each indicator is logically tied through a performance objective to a strategic goal, and ultimately to the overall mission. Finally, the group crafted data collection instruments to gather data for each indicator. The EELIAS information technology vendor converts the instruments to electronic format and connects them to the database.

Development Challenges

The EELIAS project director acknowledged that such an undertaking is a time-intensive, painstaking process. It requires significant up-front investment in both people and dollars. However, the rewards include better management, increased productivity, customer satisfaction, budget justification, and ultimately mission accomplishment. A few other factors were specifically cited as ingredients for success. These are human factors and are primarily a function of buy-in.

Management Buy-in: There is no substitute for top-level resolve combined with belief in performance measurement. Given the financial and human resources involved, the process of developing a performance system must begin with strong leadership by management. And, with management as one of the primary customers of the system, their input represents a critical part of the development process. A performance system is built from the ground-up by working-level experts, but the impetus comes from the top. EELIAS enjoyed this benefit from the outset.

Stakeholder Buy-in: A performance system touches many parties who have a great interest in the process. EELIAS' task force composition reflected this. Initially, many parties felt suspicious about what EELIAS represented. Some wondered whether a new entity was attempting to seize their power or resources. The institution of a completely transparent process, full of open dialogue, helped to ease this uncertainty. The assemblage of agency managers, OMB, congressional staff, performance experts, subject matter experts, and the grantee community, engendered a great deal of trust. Under this inclusive format, a consensus eventually emerged after groups presented their own interests and individual participants aired their disagreements. Such candid communication produced a synergy that advanced the project and eliminated barriers. All parties felt that they had something to gain by making EELIAS a useful tool.

Section III. Cross-Program Performance Measurement

The IAWG Performance Measurement Study Group was tasked to develop recommendations on performance measures for all U.S. Government-sponsored international exchanges and training programs. To the extent possible, we identified a common, unifying approach to performance measurement to assist agencies/programs in implementing an effective performance measurement framework. However, we believe that it is not possible for a useful body of performance measures (indicators) to be centrally created and applied to all international exchanges and training programs. With international exchanges and training programs as varied and diverse as the agencies that implement them, forcing a common set of indicators upon them would undermine all the benefits of sound performance measurement.

USAID, in implementing its performance measurement system, attempted to create a common basket of indicators/measures from which program managers would choose.[21] Except in areas with statistical measures that are commonly accepted worldwide (e.g., education programs: enrollment ratios, specific grade completion rates, etc.), the creation of common indicators did not work. USAID Missions need to tailor indicators to their own local and budgetary considerations. However, all Missions' strategic frameworks, including indicators, are available electronically, making it possible for Missions to borrow ideas from other Missions' indicators where appropriate.

While it is not feasible to develop a series of measures for all international exchanges and training programs, it may be possible to build upon the commonalities found among smaller groupings of these programs in order to develop similarly tailored approaches or similar measures for programs with common or related goals, objectives, and delivery mechanisms.[22] The National Academy for Public Administration's (NAPA) Performance Consortium points out that while the Results Act does not require performance measurement across similar programs,

[21] These were not directed specifically at USAID's training programs, because training is considered one among many inputs contributing to the achievement of overarching goals.
[22] Programs that can be grouped for this purpose are also commonly referred to as cross-cutting programs.

"its conceptual framework and motivating spirit move in that direction." The Performance Consortium presents the following guidance for improving performance across programs:

- List related programs and prioritize them for possible attention based on a review of likely costs and benefits.

- Look for "win/win" opportunities to demonstrate the value of programs in relation to each other.

- Examine program missions at high and broad levels, focusing especially on goals; then proceed to increasingly more specific levels.

- Use a logic model to demonstrate the relationships among programs, and present the logic model in graphic form.

- Identify both common and unique contributions made by each program to related goals. (This aspect also provides the benefit of showing program complementarity and enables organizations to self-monitor for programming duplication.)

- Explore existing data systems with an eye to using data in creative new ways to reduce costs.

- Define categories of use appropriate for the measure chosen.[23]

EXAMPLE 4

The National Academy of Public Administration (NAPA) divides goals into several broad categories. Exchanges and training activities could each fit into more than one of these categories, which can be used to provide a structure to group programs, activities, and/or measures.

- Policy -- develops approaches and methods
- Task accomplishment -- fulfills specific program goals
- Compliance -- obtains the adherence of organizations/people, e.g., regulations
- Deterrence -- stops/prevents some activity
- Customer Service -- provides assistance to someone

Center for Improving Government Performance, Helpful Practices in Improving Government Performance: An Overview of Helpful Practices, p. 10.

This guidance provides a logical framework for the IAWG to address performance measurement among smaller groups of international exchanges and training programs. Several approaches can be used to group programs so that common goals and indicators can be developed.

- **Agency/Organization:** While this approach to grouping allows linkages to agency strategic plans, it is less useful for international exchanges and training programs, many of which receive funding through interagency transfers and relate only tangentially to the administering agency's strategic plan.

[23]Center for Improving Government Performance, *Helpful Practices in Improving Government Performance: An Overview of Helpful Practices*, Focus Paper, National Academy of Public Administration (NAPA), Washington, D.C., June 1998, pp 6-7.

- **Funding Source:** Programs funded from the same sources (e.g., the Freedom Support Act) all have specific criteria and goals associated with the particular source of funds. These commonalities could possibly be tapped to develop a useful family of indicators.

- **National Interest:** All foreign policy activities of the U.S. Government must fall under at least one of the seven national interests and 16 strategic goals outlined in the 1999 International Affairs Strategic Plan. This grouping could be useful when developing end outcome goals and indicators.

- **Delivery Mechanism:** This is an appropriate and straightforward grouping for output and intermediate outcome measures. Programs with the same delivery mechanisms (such as train-the-trainer seminars, distance education programs, and academic degree programs) will have identical or similar outputs and intermediate outcome indicators.

Section IV: Conclusion

Common Challenges

Performance measurement has proven a challenge to federal government organizations. This challenge has been and continues to be recognized by Congress, OMB, and other interested entities. No single approach is going to fit all organizations. Each organization's experience with performance measurement and its ability to implement a sound system is affected by a wide range of factors, including, but not limited to:

- Institutional support for performance measurement (at all levels).
- Availability of funds for performance measurement.
- Availability of data management systems.
- Specific goals of the organization's programs.
- Delivery mechanisms.
- Time-line for achieving results.
- Support for efforts by external stakeholders (including program partners and decision makers in Congress and OMB).

With limited resources, managers often must set priorities for performance measurement activities. Managers may need to pick and choose which program activities to measure and the types of measures (indicators) to use. In these instances, managers should avoid taking the easy way out, such as relying solely on easy-to-retrieve data and/or focusing on solid programs with strong outcomes. Sampling similar programs provides one solution for spreading thin resources around. Another is to communicate the performance measurement approaches and activities that would be optimal for the program and then explain limitations that restrict performance measurement activities. Providing resource requests in tandem with these explanations will present resource allocators with clear associations between resources and performance measurement capabilities. As a benchmark, USAID informs its managers that three to ten percent of the total budget for a strategic objective represents a reasonable level to spend on performance monitoring and evaluation.[24]

Motivating employees to actively participate in and support performance measurement has also proven to be a major challenge. Many government employees view results-based management as a temporary trend that will fade away with time. There is strong resistance to adopting performance measurement systems. Learning the language and methodology of performance measurement requires a significant commitment on the part of program staff. Organizations need to facilitate employee input and maintain open lines of communication, which can encourage employees to support and participate in performance measurement.

Management organizations need to provide incentives for employees. Punitive measures may be appropriate *after* a system is firmly established; otherwise these measures might alienate employees and build distrust in early stages of establishing performance measurement systems. Additionally, punitive measures would only encourage selective reporting and misrepresentation of data. Monetary incentives are rarely feasible, considering current budget constraints, though employee incentive payments or increased funding for programs would likely be effective. Non-monetary awards provide recognition, but some employees see them as empty gestures, designed to placate and cajole them into cooperating.

Increased discretion over resource allocation and program design may be one of the best methods for encouraging employees, but this is constrained by two factors: First, front-line program implementers may not be program managers, and thus would likely not be in a position to take advantage of programming authority. Second, the trend toward budget earmarks in Congress reduces the discretionary programming options among federal government organizations.

Other challenges relevant to international exchanges and training programs include, but are not limited to, the following:

- **Access to Data:** Often data on international exchanges and training programs are not maintained in Washington, but at field locations throughout the world. Collecting and analyzing this information either requires a sophisticated (and costly) data management system or large amounts of staff time. Organizations with large international exchanges and training portfolios, such as USAID, have expended the resources needed to address this issue, but smaller organizations have difficulty justifying the cost relative to the amount and importance (on a strategic level) of data collection and management.

- **Language Barriers:** The need for translators and interpreters, which increases the amount of resources (human and capital) expended, may hamper long-term follow-up and outcome assessments.

- **Cultural and Political Barriers to Getting Data:** Some cultures reject participant tracking and assessment activities as invasive and threatening. Participants may question how such data will be used and/or fear negative attention from elements within their

[24] USAID, Center for Development Information and Evaluation, *Preparing a Performance Monitoring Plan*, Performance Monitoring and Evaluation TIPS, No. 7, 1996.

country. Even in cultures with no such overt concern, maintaining even simple directory-type data may be foreign to them. Attempting to shift these attitudes and/or work around them places a huge burden on international exchanges and training programs.

- **Quantifying Results:** The results of international exchanges and training programs are especially difficult to quantify. Many program managers feel that attitudinal changes and improvements in the policy environment can not be quantified. With the aggressive cost-cutting approach Congress has taken toward international exchanges and training programs in the past, managers also fear that quantitative data will be misinterpreted and used to eliminate programs that have very positive qualitative results.

Next Steps

The IAWG recommends that managers of international exchanges and training programs utilize the primer provided to develop or refine existing performance measurement systems. Very few of these programs systematically and effectively measure and communicate results. The guidance provided here and throughout the resources noted in this report should provide managers with the basic tools they need to enhance their ability to communicate the effectiveness of their programs and tie results to resource requests.

Congress and other high-level decision makers, however, need to recognize the challenges facing all U.S. Government agencies, but especially those implementing international exchanges and training programs. Additional resources to measure performance should be provided; otherwise, managers resort to arbitrarily cutting program budgets. Additionally, Congress should provide agency managers with planning and budgeting flexibility to augment successful programs and redesign, reduce, or eliminate poor performers.

The IAWG can make two significant contributions to performance measurement among international exchanges and training programs. First, the IAWG can continue to identify and share best practices with interested organizations. While performance measures must be tailored to the specifics of each program and activity by individuals working directly with the program, seeing examples of measures used by other organizations with similar activities may help program administrators develop new and better measures. Over time, the IAWG will continue surveying U.S. Government organizations to monitor their adoption of performance measurement systems and to study the measures they implement for their exchanges and training programs.

Second, the IAWG will identify smaller groups of programs that may benefit from the cross-program performance measurement addressed in Section III. Specifically, the IAWG will determine, when appropriate, whether specific approaches to the development of program measures and examples of indicators can be shared among programs with similar funding sources, goals, objectives, and delivery mechanisms.

Kirkpatrick Scale

One possible set of performance measures appropriate for international exchanges and training programs is based upon Donald Kirkpatrick's work in evaluating training programs. Kirkpatrick identifies the following four levels to evaluate training programs:

Level 1 -- Reaction: Measures customer satisfaction.

Level 2 -- Learning: Measures the extent to which program participants change attitudes, improve knowledge, and/or increase skills.

Level 3 -- Behavior: Measures the extent to which behavioral changes occurred as a result of the program -- the observable application of information learned.

Level 4 -- Results: Measures the extent to which change in institutions/conditions occurs. Examples include: increased production, improved quality, decreased costs, reduced frequency in accidents, etc.

Kirkpatrick explains, "The four levels represent a sequence of ways to evaluate programs. Each level is important and has an impact on the next level. As you move from one level to the next, the process becomes more difficult and time-consuming, but also provides more valuable information. None of the levels should be bypassed simply to get to the level that the trainer considers the most important." [25]

Prior to its integration into the Department of State, the U.S. Information Agency (USIA) used a modified version of the Kirkpatrick scale to measure the results of its exchange and information programs. The public diplomacy sections of the Department of State still use these levels, as depicted below:

Level 1 -- Reaction: Evidence exists that information or experience provided through USIA programs has reached key contact/target audience.

Level 2 -- Learning: Key contact/target audience show evidence of learning.

Level 3 -- Behavior: Key contact/target audience member acts in ways parallel to the U.S. national interest, or contributes to more friendly, peaceful, or cooperative relations with the United States.

Level 4 -- Results: Key contact/target audience member leads his or her institution to act in ways parallel to the U.S. national interest, or contribute to more friendly, peaceful, or cooperative relations with the United States.

[25] Donald L. Kirkpatrick, *Evaluating Training Programs*, Berrett-Koehler Publishers, Inc., San Francisco, 1988, p. 19.

USIA, which did not view the system as hierarchical, realized that most of its work fell under Level 1. Recognizing that results in Levels 3 and 4 would occur less frequently, USIA emphasized measuring results over a period of time. Level 1, which implies some output measures, determines whether the product or service was delivered to the right people. The other three levels are results-oriented: Level 2 measures change in knowledge/attitude; Level 3 measures change in behavior; and Level 4 measures change in condition.[26]

[26]Some public diplomacy elements at the Department of State further interpret these Levels as immediate outcomes (Levels 1-2), intermediate outcomes (Level 3), and end outcomes (Level 4).

International Exchanges and Training Program Performance Measurement Examples

Example 1: Academic Exchange Program

The example of the Hubert H. Humphrey Fellowship Program is based on an evaluation of the program conducted by Macro International, Inc. One goal of the evaluation was to develop performance measures for the Humphrey Fellowship. The Goals, Intermediate Outcomes, and Performance Measures for the program are taken from the final evaluation report. The Target Levels and End Outcomes are not official and are supplied as examples.

Mission: The Hubert H. Humphrey Fellowship Program strengthens U.S. interaction with outstanding mid-career professionals from a wide range of countries with developmental needs by providing the Humphrey Fellows with one year of study and opportunities to develop professional expertise and leadership skills for public service.

Goal(s): The four stated goals of the Humphrey Fellowship are to:

- Update professional expertise and leadership goals.
- Broaden understanding and knowledge of developmental issues.
- Contribute to mutual understanding.
- Establish and enhance long-lasting, productive partnerships.

In many cases, there will also be unstated or unexpected goals. The sample measures capture one such example.

Inputs: Inputs are all the resources that are put into a program that include:

- Financial resources – appropriated budget, in-kind contributions, and direct cost-sharing.
- Human resources – the number, skills, and experience of staff at the Department of State, U.S. Embassies, Fulbright Binational Commissions, the Institute of International Education, U.S. universities and internship placement sites, and the program participants.
- Time – staff hours and programmatic process time. (Frequently, time is determined by assessing staff input into the Humphrey Program at the Department of State, contracting partner organizations, U.S. Embassies, and Fulbright Commissions, in terms of Full-Time Equivalent positions.)
- Program documentation and institutional history – the developed system of procedures, policies, and institutional knowledge of the program, which is an independent variable that can affect outcomes. (This input is a factor in measuring performance from an evaluation perspective but usually is not used independently as an instrument in measuring performance.)

Process: The activities for the program include the programmatic and administrative work of the State Department sponsoring office, contracting partner organizations, U.S. Embassies, and Fulbright Commissions, such as participant recruitment, selection, pre-departure orientation sessions, the academic program (classroom work), professional and cultural enrichment activities, and practical training (on-the-job experience).

Outputs: The primary output of the program is a trained Humphrey Fellow. Secondary outputs include the number of U.S. universities and foreign countries participating, the number of Fellows' course hours, the number of Fellows' internships, long-term relationships/partnerships, funding spent in local U.S. communities, publications, and training sessions for university administrators.

Example 1: Academic Exchange Program

Goal	Intermediate Outcome	Performance Measure	Target*	End Outcome
1. To update professional expertise and leadership skills	1.1 Professional development skills acquired and applied (English, computers, presentations, grant or proposal writing, leadership, management, communication) 1.2 Knowledge in field of study acquired and applied 1.3 Knowledge disseminated (papers presented at conferences, published in journals, teaching) 1.4 Changes implemented (new policies, new practices, new programs, new businesses) 1.5 Careers advanced (new responsibilities, promotions, special assignments to committees and boards)	An increase in the professional achievements of Fellows, based on knowledge and skills acquired during the Humphrey year.	65% of the Humphrey Fellows will attribute an increase in each of the outcome areas to the Humphrey Program within five years after returning from the program. 25% of the Humphrey Fellows will advance in their careers within five years after returning from the program.	An improvement in public service in developing countries demonstrated through professional skills.
2. To broaden understanding and knowledge of develop-mental issues	2.1 Returned to work in home country or worked at an international organization 2.2 Commitment to development actualized (implemented project that addresses a community need, established an NGO)	A professional contribution to the larger social order or an increase in community development, based on knowledge and understanding acquired and a public service ethic fostered during the Humphrey year.	80% of Humphrey Fellows continue to work in public service related to the development of their country or region five years after returning from the program.	To improve the economies and infrastructure of developing countries to make them more viable partners for the U.S.
3. To contribute to mutual understanding	3.1 Knowledge of U.S. acquired (U.S. practices, culture, government, policies) 3.2 Friendships maintained 3.3 U.S. views introduced to government policymaking through positions of influence 3.4 U.S. perspectives communicated to friends, family, colleagues, and the broader community	An increase in the dissemination of the U.S. perspective, based on knowledge of the U.S. gained during the Humphrey year and relationships formed with Americans during the Humphrey year.	70% of Humphrey Fellows have regular personal contact with Americans five years after returning from the program.	The development of friendly, peaceful, and sympathetic relations between the U.S. and countries around the world.
4. To establish and enhance long-lasting, productive partnerships	4.1 Relationships maintained (professional contacts, staying up-to-date in field, involvement in U.S. or international association) 4.2 Cross-national collaborations conducted (co-authored papers, engaged in research/projects, conducted business, established trade, organized regional or international conferences)	Collaboration with contacts established during the Humphrey year on professional, business, national, or international activities.	30% of Humphrey Fellows will collaborate professionally with U.S. counterparts met during the Humphrey year five years after returning from the program.	On-going sustainable partnerships and collaborations between the U.S. and other countries around the world.
Unstated	5.1 Personal attributes developed (self-confidence, independence, motivation, cultural tolerance)	An increase in the personal growth of Fellows, based on experience gained during the Humphrey year.	85% of Fellows continue or expand their leadership roles and attribute to the Humphrey Program their personal growth in one or more characteristics listed five years after completion of the program.	Long-term development of future leaders within the countries involved in the program, or in international programs reaching across borders.

*The targets, which are not official, exemplify the degree of achievement that can be expected. It is also important to note that not all participants are expected to meet all outcomes.

Example 2: International Scientific Research Exchange

International research and scientific exchanges and training programs face unique challenges in developing a performance measurement framework. Not only are the results of the programs -- both intended and unknown -- likely to be realized years after the research concludes, but external factors make it difficult to relate those outcomes to the funded research. The following hypothetical example of a performance measurement framework for this type of international exchange and training program provides a starting point:

Organizational Mission: "To promote the progress of science; to advance the national health, prosperity, and welfare..."[27]

Goals: "...important discoveries; new knowledge and techniques, both expected and unexpected..."[28]

"...results...are rapidly and readily available and feed, as appropriate, into education, policy development, or use by other federal agencies or the private sector."[29]

Objective: Increase early diagnosis of skin cancer.

Inputs: Dollars and full-time employees.

Process: Fellowship awards to international experts in the field to work on joint research with U.S. experts at U.S. facilities.

Outputs: During three-year research collaboration, six reports and two journal articles.

Intermediate Outcomes: Reports and articles; peer reviews; citations in technical literature;[30] and use of the new diagnostic method.

Performance Measures: Percentage of time reports and articles were provided on schedule; percentage of favorable peer reviews, i.e., whether current research resources are invested in promising fields where new breakthroughs are needed;[31] the number of citations in technical literature; percentage change in use of new diagnostic method; and change in timing of diagnosis.

---

[27]National Science Foundation, *GPRA Performance Report FY 1999*, March 2000, p. 7.
[28]Ibid., p. 13.
[29]Ibid.
[30]Hatry, p. 67.
[31]NAPA, *Designing Effective Performance Measures*, p. 14.

Target: One hundred percent of reports and articles provided on schedule; 80 percent of peer reviews favorable; 50 citations in technical literature within six months of publication of reports and articles; and new diagnostic method used in 10 percent of suspected cases of skin cancer within two years after research concludes.

End Outcomes: Reduction in the number of fatalities from skin cancer.

Performance Measure: Percentage reduction in the number of fatalities from skin cancer as a result of new diagnostic method.

Target: Fatalities from skin cancer reduced by 5 percent over five years following introduction of new diagnostic method.

Example 3: Training Component of a Larger Program

Several organizations that implement international exchanges and training activities do so as components of larger programs. Performance measurement may be directed at the specific training activities or at the larger program. USAID considers training a crosscutting tool that can be used in all of its strategic objective areas to contribute to achieving results. The following example from a USAID Mission overseas illustrates this point:

A Mission assisting in developing the private sector wanted to encourage micro-entrepreneurship. One barrier identified was the long and bureaucratic process required to get a private business license; applicants would have to get approval from seven offices in the capital. The staff of these offices were often unhelpful and usually found some reason to make the applicants come back the next week. Most applicants gave up out of frustration. However, the government ministry seriously recognized the need for reform. USAID provided training to its staff in customer service and in the operation of a new computer system. Other factors that were arguably more important than the USAID training inputs but also supported by the agency were reducing the number of government offices required to sign off on licenses from seven to three (and placating the four bureaucratic turf losers), amalgamating the remaining three offices into a one-stop-shopping office, designing and purchasing the computer system, and reforming the commercial laws to facilitate micro-entrepreneurship.

If we were to take this example and attempt to not only demonstrate the overarching achievement, but to recognize the role of the training program, we also would need to recognize the importance of these additional non-training inputs/factors. The following is a hypothetical performance measurement example drawn from the information above:

Mission:	Sustainable development.
Goal:	Broad-based economic growth achieved.
Objective:	To strengthen markets through the encouragement of micro-enterprise development.
Input:	Funds and human resources necessary to provide technical assistance to government ministry, implement the training program, develop training materials, and design and purchase computer system.
Output:	Training sessions for employees of state ministry, creation of reusable training materials, tailored computer application.
Intermediate Outcome:	Simplification of process to apply for and receive a private business license.
Performance Indicators:	Reduction in the number of offices needed to approve licenses and the amount of time required to obtain licenses.

Performance Target: To reduce the number of offices approving licenses from seven to three, and to reduce the time required to apply for and receive a private business license from two weeks to two days.

Result: The number of offices required to approve private business licenses was decreased from seven to three. The remaining three offices were amalgamated into a one-stop center, and the time required to receive a private business license decreased from two weeks to two hours.

This example does not include the final stage of performance measurement, which would involve quantifying end outcomes. An appropriate end outcome and measures would be:

End Outcome: Markets strengthened through the encouragement of micro-enterprise development.

Performance Indicators: Increased number of micro-entrepreneurs/increased number of private business licenses registered over previous year; number of micro-enterprises and private businesses that remain in business three years after registration; improved overarching economic performance after three years [as measured by the Gross National Product, Gross Domestic Product, or another appropriate economic indicator].

Example 4: Field-based Training

The following example depicts a training program as a component of an overall law enforcement development effort that would likely have additional goals and training activities. The mission addresses the overall program, but only one goal -- and its corresponding performance measures -- is included. A special aspect of this program is the "train-the-trainer" component, in which the U.S. Government program attempts to create a sustainable training capacity in the host country. This is a hypothetical composite of several Department of Justice training programs.

Program Mission: To assist in transitioning existing police forces to a professional civilian police institution based on democratic principles.

Goal: To improve the criminal investigative capacity of the police.

Objectives: To strengthen basic investigative skills, promote respect for human dignity, and train trainers to teach basic investigations.

Inputs: Funds (USG, host country, etc.); human resources (instructors, course material developers, etc.).

Process: Hire U.S. law enforcement experts to write the course work with host-country experts and tailor it to local laws and customs; assist host country in selecting trainees and delivering the training; design and implement follow-on activities to facilitate the application of training, including fostering of host-country support.

Outputs: Number of police investigators trained; number of country instructors taught to teach new investigation course; new/revised criminal investigation curriculum.[32]

Intermediate Outcome: Increased understanding of investigative skills taught and how to maintain respect for human dignity during the investigative process.

Performance Indicator: Skill test administered before and after training.

Performance Target: Eighty percent of trainees improve score on skills test.

Intermediate Outcome: Creation of indigenous capacity for investigative skills training.

[32]For this example, the revised criminal investigations curriculum is an output. It could be considered an intermediate outcome, though traditionally the *use* of the curriculum would constitute the outcome.

Performance Indicator:	USG-trained in-country trainers offer training seminars through the local training academy following the conclusion of their USG-sponsored training course.
Performance Target:	An additional four seminars offered by indigenous trainers in the six months following the program.
Performance Indicator:	The percent of criminal investigators who have taken the basic investigations course.
Performance Target:	One-hundred percent of criminal investigators are required to take the basic investigation course before assuming investigative responsibilities.
Intermediate Outcome:	Police investigators use modern investigative techniques that incorporate respect for human dignity.
Performance Indicators:	Significant reduction in the use of forced confessions; number of crime scenes adequately protected; number of interviews conducted and documented; number of reports complete and timely.
End Outcome:	Improved investigation of criminal cases.
Performance Indicator:	Decrease in the number of cases refused by the prosecutor due to an inadequate police investigation.

OMB Circular A-11 (July 2000)

SECTION 200 -- OVERVIEW OF STRATEGIC PLANS, ANNUAL PERFORMANCE PLANS, AND ANNUAL PROGRAM PERFORMANCE REPORTS

200.2 Definitions.

General goal: Included in a strategic plan, this goal defines how an agency will carry out its mission over a period of time. The goal is expressed in a manner which allows a future assessment to be made of whether the goal was or is being achieved. The goal may be of a programmatic, policy, or managemental [sic] nature. General goals are predominately outcome-type goals.

General objective: Included in a strategic plan, the objective(s) are paired with a general goal and can be used to help assess whether a general goal was or is being achieved. An objective usually describes a more specific level of achievement than a general goal.

Outcome goal: A description of the intended result, effect, or consequence that will occur from carrying out a program or activity.

Output goal: A description of the level of activity or effort that will be produced or provided over a period of time or by a specified date, including a description of the characteristics and attributes (e.g., timeliness) established as standards in the course of conducting the activity or effort.

Performance goal: Included in the annual performance plan. A target level of performance expressed as a tangible, measurable objective, against which actual achievement can be compared, including a goal expressed as a quantitative standard, value, or rate. Performance goals can be either outcome or output goals.

Performance indicator: A particular value or characteristic used to measure output or outcome. Performance indicators are associated with performance goals in the annual performance plan.

Performance measure: A performance goal or performance indicator.

Program activity: A specific activity or project as listed in the program and financing schedules of the annual budget of the United States Government. (See also section 220.9 on program activity and section 82.3 on the program and financing schedule.)

Program evaluation: An assessment, through objective measurement and systematic analysis, of the manner and extent to which federal programs achieve intended objectives.

Performance Measurement Bibliography

In preparation for this report, the IAWG Performance Measurement Study Group reviewed the resources listed below. This is by no means an exhaustive list of performance measurement resources. We divided the list into two sections. The first section includes those publications cited in our report that we believe to be most pertinent to administrators of international exchanges and training programs. Within this section, the Study Group notes that Hatry's *Performance Measurement: Getting Results*, while not specific to international programs, proved an especially useful and thorough resource. The USAID TIPS sheets included here provide an excellent example of the guidance an agency can issue to inform and enlighten a wide, agency-oriented user community. The second section includes selected congressional correspondence and GAO reports on performance measurement topics. We did not include reviews or critiques of agency-specific performance plans.

Principal Sources

Hatry, Harry P., *Performance Measurement: Getting Results*, The Urban Institute Press, Washington, D.C., 1999.

Center for Improving Government Performance, *Using Performance Data to Improve Program Effectiveness*, Focus Paper, National Academy of Public Administration, Washington, D.C., June 1999.

______, *Designing Effective Performance Measures*, Focus Paper, National Academy of Public Administration, Washington, D.C., June 1999.

______, *Using Practical Program Evaluations*, Focus Paper, National Academy of Public Administration, Washington, D.C., June 1999.

______, *An Overview of Helpful Practices*, Focus Paper, National Academy of Public Administration, Washington, D.C., June 1998.

Office of Management and Budget (OMB), *Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports*, Circular No. A-11, Part 2, Washington, D.C., July 2000.

U.S. Agency for International Development's Center for Development Information and Evaluation, "Building a Results Framework," Performance Monitoring and Evaluation TIPS, No. 13, Washington, D.C., 2000.

______, "Establishing Performance Targets," Performance Monitoring and Evaluation TIPS, No. 8, Washington, D.C., 1996.

______, "Guidelines for Indicator and Data Quality," Performance Monitoring and Evaluation TIPS, No. 12, Washington, D.C., 1998.

______, "Preparing a Performance Monitoring Plan," Performance Monitoring and Evaluation TIPS, No. 7, Washington, D.C., 1996.

______, "Selecting Performance Indicators," Performance Monitoring and Evaluation TIPS, No. 6, Washington, D.C., 1996.

Additional Resources

Congressional Leadership Letter to OMB regarding expectations and views of agencies' FY 1999 performance reports due 3/31/00, July 1, 1999.

Congressional Leadership Letter to OMB regarding GAO assessment of agencies' FY 1999 performance reports, September 10, 1999.

Kirkpatrick, Donald L., *Evaluating Training Programs*, Berrett-Koehler Publishers, Inc., San Francisco, 1988.

National Partnership for Reinventing Government (Kathleen Monahan, Project Leader), *Balancing Measures: Best Practices in Performance Measurement*, August 1999. Available at http://www.npr.gov/library/papers/bkgrd/balmeasure.html

U.S. General Accounting Office, *Agency Performance Plans: Examples of Practices That Can Improve Usefulness to Decisionmakers*, GGD/AIMD-99-69, Washington, D.C., February 1999.

______, *Executive Guide: Effectively Implementing the Government Performance and Results Act*, GGD-96-118, Washington, D.C., June 1996.

______, *Major Management Challenges and Program Risks: A Government-wide Perspective*, OCG-99-1, Washington, D.C., January 1999.

______, *Managing for Results: Measuring Program Results That Are Under Limited Federal Control*, GGD-99-16, Washington, D.C., December 1998.

______, *Managing for Results: Opportunities for Continued Improvements in Agency Performance Plans*, GGD/AIMD-99-215, Washington, D.C., July 1999.

______, *Performance Budgeting: Fiscal Year 2000 Progress in Linking Plans with Budgets*, AIMD-99-239R, Washington, D.C., July 30, 1999.

______, *Performance Budgeting: Initial Agency Experiences Provide a Foundation to Assess Future Directions*, T-AIMD/GGD-99-216, Washington, D.C., July 1999.

______, *Performance Budgeting: Initial Experiences Under the Results Act in Linking Plans With Budgets*, AIMD/GGD-99-67, Washington, D.C., April 1999.

______, *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GGD-99-139, Washington, D.C., July 1999.