

## OCCASIONAL PAPERS

# Leadership Can Be Learned, But How Is It Measured?

How does leadership development contribute to measurable changes in organizational performance, that is, behavior changes in teams and the results they produce that contribute to the organization's overall goals? How should we explore and document the performance factors and experiences of participants to understand the influences of a leadership program? To answer these questions, Management Sciences for Health developed a comprehensive yet practical approach for evaluating the outcomes of its leadership development programs with health sector participants in developing countries. This evaluation involves both tracking and documenting concrete evidence of change and incorporating the human face of these data to understand the struggles and victories behind the results achieved. This paper describes the evaluation methodology used and shares the lessons learned through the experience of the Management and Leadership (M&L) and Leadership, Management and Sustainability (LMS) programs.

Effective management and leadership are critical aspects of well-performing health organizations and programs. Even the availability of technically competent health professionals, adequate facilities and supplies, and well-designed programs does not assure high performance in delivery of family planning or other health services. Good leadership and management practices are harder to assess than other aspects of high-performing health systems but are nonetheless equally important. One of our major measurement challenges, therefore, is: How do we effectively measure leadership outcomes to demonstrate the tangible benefits of investing in leadership and management which, in turn, will heighten credibility of leadership development as a public health intervention?

Authors: *Nancy LeMay* is Senior Program Officer for Monitoring, Evaluation and Communications in the Leadership, Management and Sustainability (LMS) Program at Management Sciences for Health (MSH). *Alison Ellis* is the Team Leader for Monitoring, Evaluation and Communications of the same program at MSH.

The Management and Leadership (M&L) Program (2000–05) and current Leadership, Management and Sustainability (LMS) Program (2005–10) implemented by Management Sciences for Health (MSH) were charged with measuring and documenting the main outcomes of their leadership development programs (LDPs) carried out with participants in the health sector of developing countries. This paper describes the primary outcomes of our face-to-face and Web-based LDPs and explains how to measure these outcomes, with what methods and for what purpose. The lessons learned and conclusions are drawn from our experiences as evaluators for the M&L and LMS programs over the past eight years.



**USAID**  
FROM THE AMERICAN PEOPLE

This paper is based on a chapter prepared by the Monitoring and Evaluation (M&E) Unit of the M&L Program and published by the Center for Creative Leadership in *The Handbook of Leadership Development Evaluation* (Hannum et al. 2007). The information presented here is a revision and update of the issues and lessons in that chapter. Our intended audience includes M&E specialists and evaluators interested in measurement in general, program managers who may want to know how rigorous one can be about measuring leadership, other agencies providing technical assistance, and donors.

### MSH LEADERSHIP DEVELOPMENT PROGRAMS

Since 2000, MSH has carried out face-to-face and Web-based programs in 49 developing countries to build the leadership and management capacities of health managers and strengthen the management systems required to deliver quality health services in a sustainable fashion. The primary clients of these programs are managers and their teams from Ministries of Health and nongovernmental organizations (NGOs), including private voluntary organizations, faith-based organizations, and community-based organizations working in the health sector. Another type of leadership development implemented by MSH addresses governance and transparency issues for multi-sectoral participants such as human resource management teams, Country Coordinating Mechanism teams overseeing grants from the Global Fund to Fight AIDS, TB and Malaria, or district health committees. In all cases, the unifying theme of our leadership development is helping a team, organization, or multisectoral body achieve its own goals no matter at what level of the health system they are working.<sup>1</sup>

By **leadership development program (LDP)**, we mean a structured process in which health managers and their teams learn to apply leading and managing practices (see Figure 1) to address the primary challenges they face in their unit or organization while receiving feedback and support from program facilitators. Together the team members create a shared vision, select their desired results, and decide how they will work to achieve the results. The manager's role is to support the team in making a commitment to the results and to provide the direction to make needed changes. The LDP not only teaches participants new skills for leading and managing, but also shows them a new way of working together that involves self-reflection and openness to learning.

MSH applied and refined the LDP in varied settings in the developing world and is now "mainstreaming" the capacity to deliver selected products, including the LDP. This capacity-building involves integrating the face-to-face LDP into the programs of other cooperating agencies of the US Agency for International Development (USAID), local organizations, and consultants. For example, the ACQUIRE project managed by EngenderHealth and the Tanzanian Ministry of Health is replicating the LDP with teams from new facilities in Kigoma, the province where MSH piloted the integration of the LDP approach into a family planning program supported by ACQUIRE. In Zanzibar, staff from an MSH partner organization facilitated the LDP with members of district AIDS committees. The mandate to mainstream brings a new challenge to the table: How do we effectively build the capacity of other organizations to measure and document leadership outcomes?

---

<sup>1</sup> In Nicaragua MSH has successfully used its approach to leadership development with nonhealth ministries, including the Ministry of Education and Ministry of the Family. This Occasional Paper addresses MSH's experience with measuring leadership development in the *health sector*.

Figure 1. Leading and Managing Framework



**scanning**

- identify client and stakeholder needs and priorities
- recognize trends, opportunities, and risks that affect the organization
- look for best practices
- identify staff capacities and constraints
- know yourself, your staff, and your organization — values, strengths, and weaknesses

**ORGANIZATIONAL OUTCOME**

Managers have up-to-date, valid knowledge of their clients, the organization, and its context; they know how their behavior affects others



**focusing**

- articulate the organization’s mission and strategy
- identify critical challenges
- link goals with the overall organizational strategy
- determine key priorities for action
- create a common picture of desired results

**ORGANIZATIONAL OUTCOME**

Organization’s work is directed by well-defined mission, strategy, and priorities



**aligning/  
mobilizing**

- ensure congruence of values, mission, strategy, structure, systems, and daily actions
- facilitate teamwork
- unite key stakeholders around an inspiring vision
- link goals with rewards and recognition
- enlist stakeholders to commit resources

**ORGANIZATIONAL OUTCOME**

Internal and external stakeholders understand and support the organization’ goals and have mobilized resources to reach these goals



**inspiring**

- match deeds to words
- demonstrate honesty in interactions
- show trust and confidence in staff, acknowledge the contributions others
- provide staff with challenges, feedback, and support
- be a model of creativity, innovation, and learning

**ORGANIZATIONAL OUTCOME**

Organization displays a climate of continuous learning and staff show commitment, even when setbacks occur



**planning**

- set short-term organizational goals and performance objectives
- develop multi-year and annual plans
- allocate adequate resources (money, people, and materials)
- anticipate and reduce risks

**ORGANIZATIONAL OUTCOME**

Organization has defined results, assigned resources, and an operational plan



**organizing**

- ensure a structure that provides accountability and delineates authority
- ensure that systems for human resource management, finance, logistics, quality assurance, operations, information, and marketing effectively support the plan
- strengthen work processes to implement the plan
- align staff capacities with planned activities

**ORGANIZATIONAL OUTCOME**

Organization has functional structures, systems, and processes for efficient operations; staff are organized and aware of job responsibilities and expectations



**implementing**

- integrate systems and coordinate work flow
- balance competing demands
- routinely use data for decision-making
- coordinate activities with other programs and sectors
- adjust plans and resources as circumstances change

**ORGANIZATIONAL OUTCOME**

Activities are carried out efficiently, effectively, and responsively



**monitoring and  
evaluating**

- monitor and reflect on progress against plans
- provide feedback
- identify needed changes
- improve work processes, procedures, and tools

**ORGANIZATIONAL OUTCOME**

Organization continuously updates information about the status of achievements and results, and applies ongoing learning and knowledge

**Leadership outcomes** depend on the context in which the LDP is delivered and the desired results selected by participant teams. They range from **intermediate outcomes** such as changes in work climate or in management systems and processes (for example, improving the recruitment, development, retention, and productivity of health professionals by strengthening the human resource management system) to **long-term outcomes** such as service delivery results (for example, improved quality of care, increased utilization of services, or better client satisfaction).

#### OUR APPROACH TO LEADERSHIP EVALUATION: HOW DOES IT MEASURE UP?

Our program philosophy states that leadership, management, and organizational development are not ends in themselves. They are means to improve the delivery of health services. In this paper we show that it is possible to measure leadership outcomes at two levels: the behavior changes in participating teams and the results these teams produce that contribute to their organization's overall goals. This measurement goes beyond the common practice of collecting participant feedback using a post-program evaluation form that rates the quality of the leadership training. It also differs from programs that are evaluated in terms of individual learning and changes in practice. Some base their assessment on indicators of individual improvement such as "enhanced professional knowledge." Others attempt to measure the extent to which individual changes lead to farther-reaching outcomes, including organizational, system, and societal changes.

Our approach to measurement developed through a process of trial and error and experimentation with a variety of methods. Over time we have expanded our reporting focus from primarily tracking activity completion to measuring changes in client organizations or in

the populations served by these organizations. We were inspired by the work of the Population Leadership Program (PLP) and the Sustainable Leadership Initiative (SLI) at the Public Health Institute<sup>2</sup> and the Leadership Evaluation Advisory Group (LEAG), all based in Oakland, California.<sup>3</sup> While also focused to a certain extent on individual leadership skills and personal transformation, PLP developed the EvaluLEAD framework for evaluating leadership outcomes at the individual, organizational, and community levels using a combination of quantitative and qualitative approaches (Grove 2002). We found this tactic extremely useful for our particular application. Given the mandate of LMS to mainstream and scale up leadership development, we are still learning how best to measure the expansion of our leadership programs and how best to train other organizations to monitor and evaluate leadership outcomes.

---

<sup>2</sup> <http://www.phi.org/>

<sup>3</sup> <http://www.leadershiplearning.org/author/leadership-evaluation-advisory-group-members>

## SECTION 1: How Do We Develop Managers Who Lead?

We define a “manager who leads” as someone, at any level of an organization, who applies the leading practices shown in Figure 1.

### THE LEADING AND MANAGING PROCESS

Managers and their teams from all levels of an organization are invited to participate in the LDP. The **focus on teams** rather than individuals is a unique feature of the leadership program. Findings from evaluations conducted under the M&L Program indicate that the composition of the team can affect its ability to address the challenge selected and to achieve desired results. Intact teams seem to have the best success in the long run, not surprisingly. An intact team is one that has worked together on a regular basis prior to the program and continues to work together after the program ends. Members of an intact team who go through the leadership development process and who continue working together are also more likely to be affected by the changes in social norms within the group. More effective work habits emerge and are mutually reinforced, thereby sustaining the changes within the team.

Another key ingredient is the premise that **leadership can be learned**. Using an action-learning approach, participants learn to apply a set of leading and managing practices to address their real workplace challenges over time (MSH 2005a). Whereas traditional leadership programs often physically and psychologically separate the participant from his or her work environment, the LDP connects the training to current challenges facing participants in their workplaces.

Our approach also assumes that leaders are not born but developed: leading and managing are skills that can be learned. Furthermore, these

skills are essential to achieving results. Our leadership programs maintain that all members of a team, regardless of their positions, can learn to lead and manage well in order to create a positive work climate and achieve results. For the purposes of the LDP, **leading well** means enabling others to face challenges, achieve results, and create the positive future that people envision. **Managing well** means ensuring that sound strategies and approaches are in place and resources are used effectively.

The LDP approach draws upon the work of experts in leading change, including Kotter (1996) and Heifetz and Laurie (1997). Participants’ feedback has also shaped the program. Feedback from participants’ evaluations and formal program evaluations has influenced the terms we use to discuss leadership as well as the length and format of our training and some of the training tools.

### A FOCUS ON CHALLENGES

The centerpiece of the leadership development approach is the Challenge Model (MSH 2005a), a simple yet powerful analysis tool derived from the Performance Improvement process (Luoma and Voltero 2002). (See Figure 2.) Participating teams use the model to create a shared vision of success, analyze their local conditions, and select a challenge from their workplace that is important and compelling. All teams are encouraged to select a challenge in a specific technical area, such as family planning/reproductive health, child health, HIV& AIDS, or human resource management.

The team then carries out the following steps:

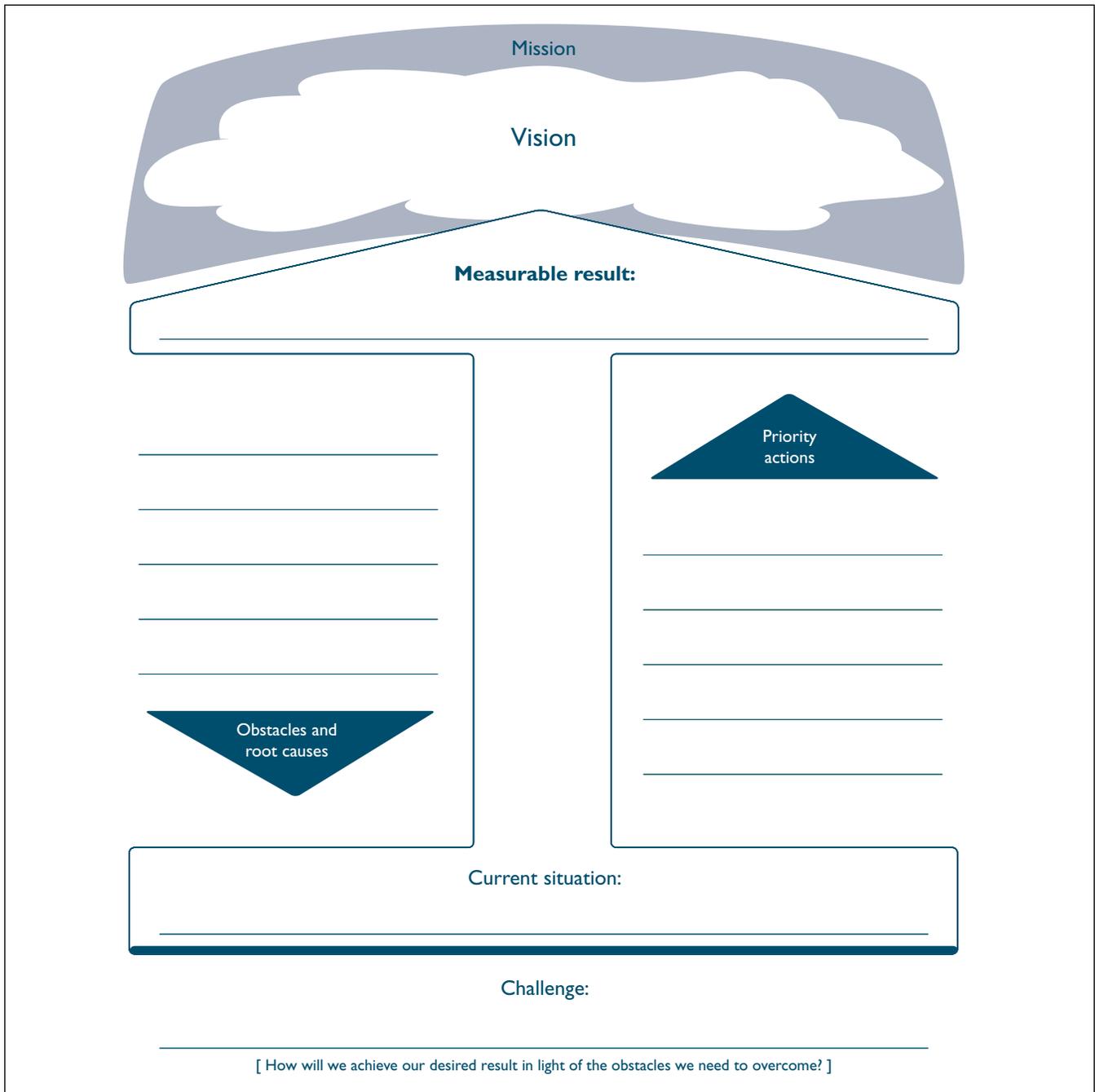
- measures the baseline in relation to the challenge;

- agrees on its expected measurable result;
- identifies main obstacles and their root causes that must be addressed to achieve the desired result;
- selects a series of priority actions to address the challenge;

- defines indicators to measure its results.

This information is translated into an action plan that serves as a management tool for participating teams and forms one basis for evaluating the outcomes.

Figure 2. The Challenge Model



## STRATEGIES FOR CONDUCTING LEADERSHIP DEVELOPMENT PROGRAMS

MSH facilitates leadership programs through face-to-face (on-site) and virtual (Web-based) mechanisms. The face-to-face program is delivered to teams who most often come from a single organization or health program,<sup>4</sup> whereas the virtual program is delivered to teams from one or more organizations in a single country or region.

**Face-to-face leadership development programs.** The face-to-face leadership development program lasts from four to nine months and is intended for members of preexisting teams from the public sector or private, non-governmental, or community organizations. During the program, participants attend a series of workshops facilitated by international or local MSH staff or consultants and then complete follow-up assignments together at their workplaces. The program engages a cofacilitator or coach from the client organization to provide ongoing support to teams. Between the workshops, the team members organize meetings during which they discuss what they have learned and review their progress in preparing or implementing their action plans. Through this process they learn to apply specific leading and managing practices to address their selected challenges and achieve their desired results. Outcomes are defined in the action plans the teams produce during the program. They receive regular coaching throughout the program by program facilitators (face-to-face or long-distance using e-mail) and local health managers.

---

<sup>4</sup> Recent exceptions are the series of face-to-face LDPs in Guyana delivered to several different organizations participating in the Guyana HIV & AIDS Reduction and Prevention (GHARP) Project, and the LDP in Nepal, which involved more than 25 teams from multiple sectors, including health, women in development, forestry, and agriculture.

## The Virtual Leadership Development

**Program.** The Virtual Leadership Development Program (VLDP) is a 12- to 16-week blended learning program. Blended learning means the program includes Web-based facilitation, course material, and individual exercises combined with face-to-face (on-site) team meetings for reflection and shared learning among participants. Participants from public and private organizations enroll as teams in the VLDP.

During each module, participants carry out individual work on the VLDP website, which includes reading modules, case studies, and editorials, completing exercises, and participating in electronic discussions. Following this, participants convene for face-to-face meetings with other team members at their workplaces to discuss what they learned during the module and to complete group work. As they move through the online modules, teams can communicate with each other using two of the website features: the Café and the Forum. The Café is a location on the VLDP website where participants are encouraged to exchange ideas on themes presented in a threaded discussion format. The Forum is another location on the website, where at the conclusion of each module a coordinator from each team describes how the group members worked together on the assignment for the module and what they produced.

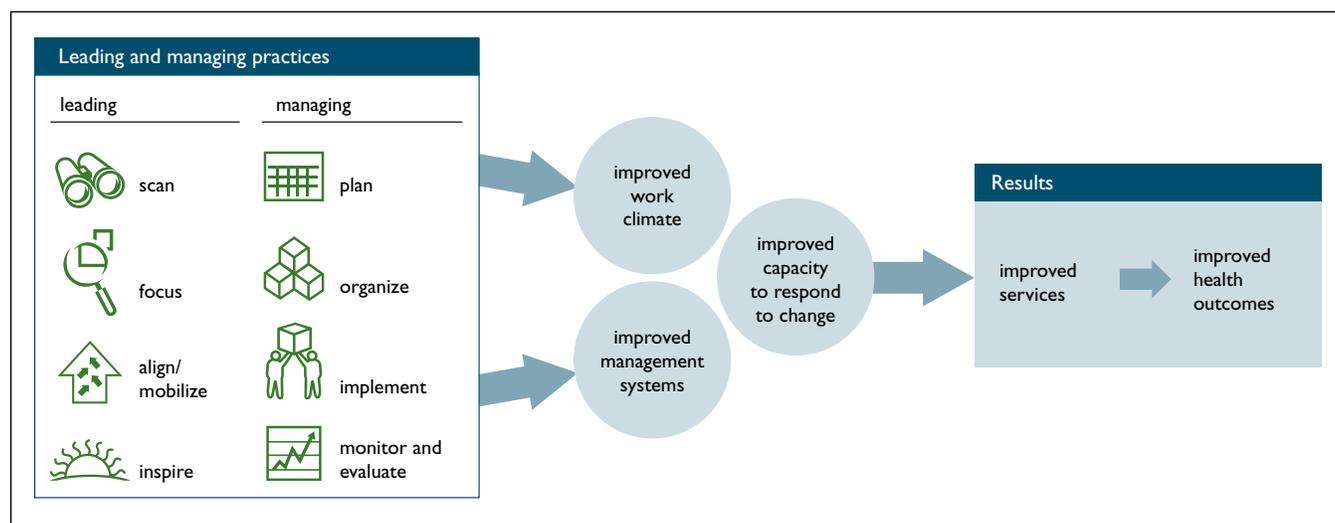
The program content consists of seven modules on topics such as leadership in health institutions, how to address a leadership challenge, leadership competencies, communication, and change management. As in the face-to-face LDP, teams in the VLDP produce action plans defining their expected measurable results, priority actions, and indicators. They also receive regular coaching and feedback from program facilitators during the program via e-mail and postings on the program website.

## SECTION II: Making the Case for Measuring Leadership

The assumption that leadership can be learned supports the idea that leadership outcomes can be measured. Each of the elements of leadership development—learning materials, workshops, online modules, feedback and support, and assistance to build specific skills, change behaviors, and work toward concrete results—can be monitored and documented. The results model in Figure 3 illustrates the program principle that measurement of leadership and management

capacity is not an end in itself; rather, working on leadership and management is a means of improving work climate and management systems and eventually of strengthening health services. Our evaluations focus on results at the output and outcome levels and the process used to achieve them. As in the open systems perspective supported by the EvaluLEAD framework, our aim is to determine contribution to results rather than demonstrate causality.

Figure 3. The Leading and Managing for Results Model



### LEADERSHIP OUTPUTS

Leadership outputs are defined as the application of specific leading and managing practices (scan, focus, align/mobilize, and inspire) by participating teams both during and after the leadership program. Managers and team members at any level of an organization can use these practices to address workplace challenges. The ability of the team members to apply the leadership practices in their workplace is the direct product (output) of the leadership development process.

MSH developed a set of indicators as a self-assessment tool for teams to monitor their behavior with respect to the four leading practices both during and after an LDP (see Table 1). A score sheet of five to seven items is used to assess the fulfillment of each indicator (see the example in Table 2 for the scanning indicator). Only a few of our leadership programs to date have adopted or applied the indicators, since it is usually premature to use them during a leadership program when teams are just learning to apply the four practices. They are most useful in assessing a team's performance before

the program to determine needs for leadership development or as a follow-up measure of a team's ability to continue applying the practices after completion of the LDP or VLDP.

To gain insights on behavior change among program participants and the distinct processes the teams used to achieve their results, we often collect information on the use of the leading practices through post-program questionnaires, interviews, and focus group discussion with participants and nonparticipants, including

those who report directly to the team manager and those who do not. While we have not systematically tracked the leadership indicators, they still provide a valuable summary of the key actions that make up each leadership practice, which the evaluator should probe during individual and group interviews. We recognize that participant observation would add greater depth to our inquiry on the changes in norms and behaviors within the group, but our ability to carry out observation of practices in the workplace has been limited by time and funds.

**Table 1. Output Indicators for Leadership Development**

<i>Leading Practice</i>	<i>Indicator</i>
Scanning	The work group can provide valid and relevant evidence about the nature of its internal and external environment, the quality and extent of its performance, and the resources available on best practices and can identify challenges within and facing the team.
Focusing	The work group has identified priority challenges to be addressed within a defined time and selected measurable actions that address barriers to achieving results.
Aligning and mobilizing	Work group responsibilities and resources are internally aligned and work group goals are externally aligned in order to address selected challenges and meet stated objectives.
Inspiring	Work groups are committed to the organization's mission and to continuous learning, improvement, and innovation.

## INTERMEDIATE OUTCOMES

We track two different intermediate outcomes defined at the team level: **better work climate** and **changes in management systems and processes**. These are considered to be the necessary precursors to improving health services (long-term organizational outcome).

Work climate is an indication of a team's internal well-being and growth. It is defined as the prevailing workplace atmosphere that is experienced by the members of a given work group.

Climate is what it feels like to work together in a group (MSH 2002). Every organization, office, and team has a climate that affects how people behave at work, which, in turn, influences workers' motivation. According to the Leading and Managing for Results Model, work climate is an intermediate outcome that is sensitive to change as a result of a leadership development process involving managers and their teams. This relationship involves two assumptions: (1) when team members work together successfully on a shared challenge, a positive work

climate is created and (2) work climate influences the achievement of long-term outcomes. Therefore it is a suitable intermediate outcome along the continuum of results. Research from the business and education sectors has shown that teams with a positive, supportive climate

tend to perform well and achieve their desired results (Goleman 2000; Laschinger, Finegan, and Shamian 2001). Better-performing teams contribute to better organizational performance, which translates into improved health service delivery.

**Table 2. Scanning Score Sheet with Key Questions**

In the past three months has the work group . . .			
<i>Item</i>	<i>Yes</i>	<i>No</i>	<i>Evidence (If Yes)</i>
1. Reviewed and discussed information from the following sources:			
➤ Service statistics?			
➤ Client satisfaction surveys?			
➤ Community-based surveys?			
➤ Policy-related issues or statements?			
➤ Other, please specify _____			
2. Reviewed and discussed the strengths, weaknesses, and/or needs of the work group members?			
3. Exchanged ideas or information with external or internal partners?			
4. Requested input and participation from all work group members during regular meetings?			
5. Reviewed best practices related to work group goals and objectives?			

Because work group climate cannot be observed directly, it is estimated through the use of a questionnaire measuring the perceptions of team members. In 2002, MSH developed a brief assessment form called the Work Group Climate Assessment (WCA), a self-scored questionnaire. The items in the survey (see Table 3) are derived from the work of George Litwin and Robert

Stringer, who developed the first surveys to measure climate in corporate environments (Litwin and Stringer 1968; Stringer 2002). MSH validated the WCA in 2004 with 42 teams from different administrative levels in Brazil, Mozambique, and Guinea (Perry et al. 2005). The Stringer Organizational Climate Survey was used as the standard during the validation study.

**Table 3. Items from the Work Group Climate Assessment**

<i>I feel that in my work group . . .</i>	<i>How are things now in your work group?</i>				
	<i>Please rate each item on a scale from 1 to 5 where:</i>				
	<b>1</b> = not at all <b>2</b> = to a small degree <b>3</b> = to a moderate degree <b>4</b> = to a great degree <b>5</b> = to a very great degree				
1. We feel our work is important.	1	2	3	4	5
2. We strive to achieve successful outcomes.	1	2	3	4	5
3. We pay attention to how well we are working together.	1	2	3	4	5
4. We understand the relevance of the job of each member in our group.	1	2	3	4	5
5. We have a plan that guides our activities.	1	2	3	4	5
6. We understand each other’s capabilities.	1	2	3	4	5
7. We seek to understand the needs of our clients.	1	2	3	4	5
8. We take pride in our work.	1	2	3	4	5

The WCA is meant to be used before and after a leadership program to provide comparative data on the work climate within intact teams. To apply the questionnaire, all team members (both managerial and staff) respond. Respondents rate how they feel about each item using a Likert scale of 1 to 5. The scores are then tabulated across all items to produce individual-level composite scores and an overall climate score for the team. Using these scores, comparisons can be made between work groups in an organization, between pre- and post-intervention assessments of the same work group, or between a single work group and a predetermined value of climate serving as a target goal.

The secondary purpose of the WCA is to engage team members in a conversation about their climate so that together they can find ways to improve it. Team members respond individually to the survey and afterward they are encouraged to discuss and act on the results together. The full WCA instrument and instructions (MSH 2005b) are available on MSH’s website at: <http://erc.msh.org/mainpage.cfm?file=96.9htm&module=toolkit&language=English>

Evaluating progress in strengthening a particular **management system** is often based on in-depth interviews and/or focus group discussions with the program facilitators, internal and external stakeholders, and staff of the client

organization, as well as a review of crucial documents linked to or produced by the management system under review (for example, guidelines, operational plans). In some cases MOST<sup>5</sup> is used to detect improvements in a single management system or across several management systems at a time.

## LONG-TERM OUTCOMES

The desired outcomes of the leadership program are defined by the participants themselves and depend on the workplace challenge they choose to address during the program. As the results model shows, MSH's leadership programs aim to contribute to improving health outcomes (for example, changes in the knowledge, attitudes, or practices of a target population or client group or changes in morbidity and mortality rates). We do this by developing leaders who can directly improve the delivery of health services. Normally we do not measure impact-level results (health outcomes) because these are too long term for our time and funding constraints.<sup>6</sup>

When appropriate, we do measure service delivery results. To expect changes in health services, participants in the leadership program must have a direct responsibility for health services (such as in a district health team or health facility personnel). When this is the case, the results of the leadership programs are measured

---

<sup>5</sup> The Management and Organizational Assessment Tool (MOST) is a self-assessment tool that uses a participatory approach to identify deficient areas in an organization's management systems and define how the weaknesses will be addressed. Results of a MOST assessment are usually used to orient subsequent technical assistance.

<sup>6</sup> We encourage teams in the leadership program to select challenges that can be accomplished in a short time (six to eight months), which precludes working directly toward results that show an impact on health. One exception to date is the leadership program in Aswan Governorate, Egypt, which has succeeded in reducing the maternal mortality rate over the past several years. The governorate was able to take on such a long-term challenge because they did not have to report short-term results to a donor.

in terms of changes in health service delivery (for example, changes in the number of clients served or the quality of services). The ability to measure change at the service delivery level is determined by several factors, including:

- the length of time and nature of our work with a team or organization;
- the challenge selected by participants;
- whether interventions are focused on addressing organizational challenges at the central level, district level, or local level;
- the functions, roles, and responsibilities assigned to the team involved in the leadership program (whether central, regional, or local).

Measuring services is not appropriate when teams have not selected a service delivery challenge to address during the program or when the scope or time frame of the program is too limited to accomplish results at this level. In such cases, we often measure intermediate outcomes, such as changes in work group climate or in organizational management processes and systems, to demonstrate whether or not the teams have made progress in the direction of improving health services.

## EVALUATION TACTICS

Our evaluations are summative, reviewing both process and results. The team forms the unit of analysis because it is the focus of the face-to-face and virtual leadership programs. We use a balanced mix of qualitative and quantitative methods. The choice of methods depends on the objectives of the evaluation, the types of challenges selected by the teams and their expected results, and the content of the action plan under review. That said, the quantitative methods generally include the use of the WCA to measure climate at the team level and

the collection of indicator data from teams to measure results according to their action plans. In cases where it is possible to measure changes in health services, service delivery data (usually service statistics) are provided by the teams and analyzed.<sup>7</sup> Other methods at the service level can include the use of observation checklists to assess provider performance and facility audits to assess clinical infrastructure and equipment. For virtual programs, e-mail questionnaires are sent to all teams that complete the VLDP.

The corresponding qualitative portion of the evaluation aims to understand and document changes in behaviors and processes within the teams both during and after the program. We take qualitative information seriously, to ensure that participants have a voice in our evaluations. Methods include document reviews as well as focus group discussions and semi-structured interviews with samples of program participants and nonparticipants. At the service level, methods can also include client exit interviews to measure client satisfaction. For virtual programs, telephone interviews are carried out with carefully selected informants.

All LDP and VLDP training materials as well as evaluation instruments are translated into the participants' primary language. From time to time we have encountered difficulties with translation issues, such as during the application of the WCA in Mozambique, where the majority of health staff at the peripheral level did not speak Portuguese. In this case, a WCA facilitator interpreted for respondents in their local language.

---

<sup>7</sup> While using service statistics for measuring outcomes is often expedient and economical, there are also challenges from a measurement perspective. Some of the primary challenges are outlined in "Data and Information Systems: Issues for Performance Measurement," in Perrin et al. 1999, pp. 83–131.

Due to time and funding constraints, we cannot carry out interviews with all participants in a leadership program. Therefore our evaluations most often use a purposive sample. Purposive sampling is a form of nonprobability sampling in which respondents are selected according to a specific plan or purpose. Purposive sampling is different from probability sampling, whereby each member of the population has an equal chance of being sampled and the results can be generalized to the sample population. Purposive sampling is useful for reaching a targeted sample and when sampling for proportionality is not the main concern. The disadvantage of a purposive sample is that it is hard to know how well the sample represents the population. It is important to qualify the findings and note whether people left out of the sample might behave differently from those who were selected.

For example, in our evaluations, teams are usually rated and selected according to criteria such as the quality of their action plans and their adherence to the SMART<sup>8</sup> principles for defining measurable results. These criteria are intended to ensure, as much as possible, maximum variability in the performance of teams included in the sample.

## KEY QUESTIONS WE ADDRESS

Our leadership program evaluations are generally based on a similar set of key questions. This allows us to synthesize the lessons learned across programs and evaluations. It is useful to triangulate several data sources for the same question to verify the information collected. Examples of the types of questions addressed in our evaluations include:

---

<sup>8</sup> SMART: Specific: to avoid differing interpretations; Measurable: to allow monitoring and evaluation; Appropriate: to the problem, goals, and strategies of the organization; Realistic: achievable, challenging, and meaningful; and Time bound: with a specific time period for achieving them.

- What technical assistance approaches and tools were used in delivering the LDP?
- What organizational challenges did the teams address through this program?
- What processes did participants establish to address their challenges (for example, team meetings, analyzing local data, consulting internal stakeholders)?
- Did teams develop action plans to address their challenges? If so, were all planned activities carried out? Were activities implemented that were not included in the action plan?
- How did the teams and their organizations monitor their progress in addressing their challenges?
- To what extent did the teams and their organizations achieve their expected results?
- What other results were achieved that are unrelated to addressing their challenges?
- What motivated participants to achieve their desired results? What prevented them from achieving them?
- Did the teams continue to work together to address another challenge after the leadership program ended? If so, what processes were used and how were they similar to or different from how they worked together during the LDP? What motivated their participation and commitment after the formal program ended?
- To what extent was content from the LDP shared with staff members who did not participate in it?
- In what ways did the LDP affect the performance of individuals, teams, and the organizations as a whole?

## SECTION III: Lessons Learned from an Evaluation Perspective

We learned three kinds of lessons in the course of this work: lessons about measuring long-term outcomes, lessons about measuring work climate, and lessons about evaluating the VLDP.

### WHAT WE LEARNED FROM MEASURING LONG-TERM OUTCOMES

**Lesson 1: The quality of participants' M&E systems is a critical factor in the ability to show and sustain results at the service delivery level.** The client organization's M&E system determines its ability to collect and process data and provide accurate information. The ability to track changes depends on the quality and accuracy of service statistics and, as in the case of the leadership program in Aswan, Egypt, the appropriateness of the indicators selected. Teams in the Egypt program that had selected family planning as their challenge used an indicator mandated by the Ministry of Health (MOH), which did not follow internationally accepted guidelines for calculating couple years of protection. Moreover, the indicator they used did not accurately measure the teams' family planning service results, in some cases underreporting results, in other cases overreporting them.

**Lesson 2: Working with clients to monitor progress and measure final results as an intrinsic part of a leadership program can create a "culture of information" that outlasts the program itself.** The leadership program carried out by the MOH and the M&L Program in Nampula Province, Mozambique, from 2004 to 2005 demonstrates this principle. The program used simple management and leadership tools to assist six districts and 11 health units and their communities to improve discrete aspects of service delivery. MOH and LDP facilitators supported the health units

in developing action plans with measurable indicators and feasible monitoring mechanisms. In the process the program created a "culture of information" among health care managers and providers in a setting where performance had never been monitored and M&E skills were very limited. This culture of information represented a new way of operating at the peripheral level based on the use of data for planning, verification, and reflection. The final program evaluation in 2005 showed that, for the most part, the health units achieved surprising results, given the low level of quality in the units and their extreme lack of resources. For example, health services improved in 10 of 11 health units, resulting in shorter waiting times for patients, improved cleanliness and hygiene, increased biosecurity, improved quality of food, construction of a maternity waiting home, and more accurate medical records. Furthermore, the health units were successful in measuring their performance using tools they had developed themselves. Table 4 shows the types of results that were monitored and reported by health units in this program.

Prior to the program, teams from the participating health units had minimal skills in developing indicators and monitoring performance. They learned to develop indicators that were appropriate to their challenges yet feasible to measure in their setting. For example, they had to find a way to define the criteria for cleanliness, hygiene, and biosecurity in a setting where they had few or no cleaning supplies and the health units were in a state of decay. While developing these indicators was difficult and time-consuming, participants were much more likely to take ownership of the process because they had helped define how their success would be measured.

**Table 4. Types of Leadership Results Documented by Selected Health Units in Mozambique**

<i>Name</i>	<i>Health Unit Type*</i>	<i>Challenges</i>	<i>Reported Results, Nov. 2003–Oct. 2004</i>
Meconta Health Center	Health Center Type 1	<ul style="list-style-type: none"> <li>■ Increase percentage of births in the health facility from 25% to 35%</li> <li>■ Increase availability of water from 2 to 12 hours/day</li> </ul>	<ul style="list-style-type: none"> <li>■ Increased percentage of births in the health facility from 25% to 35%</li> <li>■ Completed construction of maternity waiting home in Nov. 2004</li> <li>■ Dug but did not complete well</li> </ul>
Namialo Health Center	Health Center Type 1	<ul style="list-style-type: none"> <li>■ Increase average monthly percentage of contraceptive users from 15% to 50% of quota defined by the Provincial Directorate of Health (774 users)</li> </ul>	<ul style="list-style-type: none"> <li>■ Increased average monthly percentage of contraceptive users from 15% to 38% in first 9 months of 2004</li> </ul>
Ilha de Moçambique Health Center	Health Center Type 1	<ul style="list-style-type: none"> <li>■ Improve biosecurity</li> <li>■ Improve cleanliness and hygiene</li> </ul>	<ul style="list-style-type: none"> <li>■ Increased percentage of biosecurity criteria met from 16% to 83%</li> <li>■ Increased percentage of cleanliness and hygiene criteria met from 11% to 67%</li> </ul>
Lumbo Health Center	Health Center Type 2	<ul style="list-style-type: none"> <li>■ Improve internal and external cleanliness and hygiene</li> <li>■ Improve the diet of inpatients</li> </ul>	<ul style="list-style-type: none"> <li>■ Increased percentage of cleanliness and hygiene criteria met from 20% to 74%</li> <li>■ Constructed biomedical waste container</li> <li>■ Painted the health center</li> <li>■ Built a kitchen</li> </ul>
Nacala Porto General Hospital	Regional Hospital	<ul style="list-style-type: none"> <li>■ Increase number of births in the health facility</li> <li>■ Improve quality of inpatient care</li> </ul>	<ul style="list-style-type: none"> <li>■ Increased monthly average number of births in the facility from 302 for first 6 months of 2003 to 350 for first 6 months of 2004</li> <li>■ Data not reported on inpatient care</li> </ul>
Carapira Health Center	Health Center Type 1	<ul style="list-style-type: none"> <li>■ Reduce waiting time for pediatric visits</li> <li>■ Reduce number of errors in inpatient registry and Kardex (medication registry)</li> </ul>	<ul style="list-style-type: none"> <li>■ Reduced waiting time for pediatric visits by 2.5 hours</li> <li>■ Reduced errors in the inpatient registry from 9 in Dec. 2003 to 0 in Oct. 2004</li> <li>■ Reduced errors in the Kardex from 100 in Dec. 2003 to 0 in Oct. 2004</li> </ul>

\*A Type I Health Center offers basic primary health services and provides inpatient maternity care for vaginal deliveries; it may have a medical technician on staff but does not have a physician. A Type 2 Health Center does not offer inpatient maternity services and has no technical medical staff. A regional hospital has physicians and offers a range of medical and surgical services.

A nurse at the Ilha de Moçambique Health Center explained how the new culture of results is contributing to better services: “We didn’t encourage our staff to step back from their work and evaluate their activities before this program. We had plenty of ideas, but we asked for help and money when we encountered problems and we never evaluated ourselves. Now the hospital wards have a healthy spirit of competition regarding which ward has the best results when their activities are matched against the criteria that we all established.”

M&L conducted an evaluation five months after the leadership program ended. The evaluation revealed that several teams were still using their tools to monitor progress and, importantly, they were able to chart their post-program performance compared to performance during the program.

**Lesson 3: Not all leadership results are quantifiable. A mixed method approach is the most appropriate way to measure the results of a leadership program.**

Because leadership development is as much about the process as it is about the results, evaluators should always consider using a balanced mix of qualitative and quantitative methods to collect data. Evaluations that rely solely on quantitative data such as service delivery results may conceal important changes in team dynamics and their interactions with the larger system around them. The PLP, LEAG, and SLI have been in the forefront of developing ways to evaluate leadership development in the health sector. They have promoted the importance of triangulating methods and worked to develop a conceptual guide for assessing leadership development. The result is the EvaluLEAD framework, which outlines the evocative and evidential lines of inquiry necessary to capture outcomes at the individual, organizational, and societal levels (Grove 2002 and 2005). Our use of mixed methods was inspired

by this framework.

The Mozambique evaluation provides a good example of a balanced mix of methods and data sources, including:

- reports and other documents related to the Health Systems Strengthening (HSS) Program;
- reports from the November 2004 evaluation workshop in Nampula Province on health units’ progress in implementing their action plans;
- reports on the progress of the MOH’s Annual Operational Plans for 2004 and the first trimester of 2005;
- observations during site visits in November 2004 to Carapira, Lumbo, and Meconta health units and follow-up visits in April 2005 to Carapira, Lumbo, Meconta Mossuril, and Ilha de Moçambique health units;
- results of the WCAs at baseline and after completion of the program in the Directorate of Administration, Directorate of Reproductive Health, and participating health units;
- comments from program participants in the LeaderNet<sup>9</sup> virtual forum “Achieving Results in Low-Resource Settings: Stories from Mozambique” held in June 2005;

---

<sup>9</sup> LeaderNet is a global community of practice—a group of practitioners who regularly engage in sharing and learning in a specific technical area—that was developed during the M&L Program and continues to be implemented in LMS. It operates using both face-to-face meetings and through the Web, telephone, and fax, depending on the needs of its members. LeaderNet serves as a resource for managers and other professionals seeking to improve their leadership and management skills. Participants in all LMS programs, LDPs, VLDPs, and seminars offered by the Global Exchange Network (GEN) for Reproductive Health, among other programs, are invited to join LeaderNet: <http://www1.msh.org/projects/lsm/ProgramsAndTools/LeadingAndManaging/LeaderNet.cfm>.

- semistructured interviews with:
  - selected participants in the Challenges and Quality Programs in Nampula Province;
  - staff in the health units who did not participate in the workshops;
  - facilitators of the Challenges and Quality Programs;
  - selected informants from M&L Mozambique staff and key managers in the MOH;
  - representatives of the international donor common fund (SWAP).

Another example comes from the pilot LDP in Aswan, Egypt, carried out from 2002 to 2003. This program was delivered to 10 teams of doctors and nurses at the district and health facility levels. Expected outcomes were defined as improved climate and improved family planning, prenatal, and postpartum care services. The evaluation in 2003 relied entirely on service statistics to measure outcomes. Although significant improvements were measured in many of the outcome indicators related to health services, the evaluator was not able to explain in any detail what had led to these changes.

In 2004, a year after the program in Egypt ended, MSH evaluated the replication of the same program with another set of district and health facility teams. This time the program was implemented by the Aswan Governorate itself with no financial support from M&L and with limited, long-distance technical assistance. For this evaluation, a mix of qualitative and quantitative methods was used to capture changes in outcomes. Quantitative methods were again based on an analysis of service statistics. Qualitative methods included semistructured interviews and focus group discussions with participants

and facilitators. These interviews provided insight into the processes that took place within the district and facility teams and throughout the health system during the replication of the leadership program. This second evaluation provided a much richer account of program results, with evidence of changes in services coupled with insights to help explain how the transformation in service delivery was associated with the leadership development process.

#### **Lesson 4: Carefully selecting informants and creating a “safe” environment are essential to producing high-quality data.**

When selecting key informants or focus group members, it is important to keep a few pointers in mind. First, respondents should represent the characteristics of the larger team that participated in the program. For example, if the team included doctors, nurses, counselors, and outreach workers, then the sample of respondents should include these types of personnel. Second, to promote objectivity, the program facilitators or program manager should not choose the respondents. They should be selected instead by the evaluator according to criteria established in the evaluation protocol. In addition, conducting focus group discussions with participants and nonparticipants is useful to gauge the indirect effects on leadership capacity in the two groups and confirm the accuracy of the responses from those who participated in the program report.

There are some risks associated with conducting focus group discussions and interviews at the participants’ workplaces. If a respondent discusses work relationships with his or her peers or supervisors, this information must be kept confidential. For this reason, evaluators should choose interview questions with care and make sure the setting is private. Focus groups should be homogeneous, or it may be difficult to elicit candid responses. For example, staff and subordinates should be grouped together in one focus

group or set of focus groups, and managers, directors, and supervisors should be included in a separate focus group. Finally, the evaluator needs to establish a confidential atmosphere where there is no risk of retribution to encourage open discussion and honest input. While focus groups and interviews may be recorded, the verbatim transcripts are only for the evaluator's use and should not be shared with respondents' peers or supervisors.

**Lesson 5: Teams often need assistance to develop high-quality action plans that can serve as an evaluation tool.** During the leadership program, each team selects a challenge such as improving delivery of family planning services, and develops an action plan to identify the activities needed to achieve the desired results. The team also defines indicators for measuring baseline and final results. Thus the action plan links program inputs with desired outcomes.

In theory, this process provides a sound basis for measuring outcomes. In practice, however, it can be problematic to depend on teams to supply data related to their outcomes. Sometimes the data are not available when it is time to evaluate progress. Except in programs that use service statistics for monitoring results, teams may not monitor their performance using the indicators in their action plans.

There are several reasons for this problem. In some cases, team members change after the program ends. This is especially common in the public sector. Many teams do carry out their plans but may monitor performance using organizational-level indicators rather than the indicators in their action plans. In other cases, the team does not integrate its action plan into the organization's annual plan, so there may be inadequate resources to monitor using action plan indicators. Sometimes, key stakeholders in the organization are not supportive, so

teams either do not fully carry out their action plans or are not motivated to monitor their performance. Finally, while some teams produce action plans that are written logically and clearly—with well-distributed responsibilities, well-defined indicators, SMART expected results, and appropriate data sources—in other cases, it is unclear how activities in the plan will lead to achieving the desired results.

A well-developed action plan is more likely to be implemented and its indicators used than a poorly defined plan. Our prerequisite for using an action plan as the basis for an evaluation is ensuring that the plan meets the SMART criteria. To achieve this, the leadership programs have engaged M&E Unit staff to review and provide feedback to participants on their action plans, a feature of the VLDP that is yielding increasingly better results. The purpose of the review is to check that (1) the selected challenge is immediately actionable and will contribute to the intended results, (2) the activities logically address the challenge, and (3) the proposed indicators are appropriate to measure progress. To make sure the plan meets the SMART criteria, teams are given guidance using the following questions:

- Are the mission and vision of the team clear, especially as they relate to the organization's overall mission?
- Is the challenge selected by the team clear?
- Are the desired results defined by the team clear and achievable within six to nine months?
- Are activities logically related to the desired results?
- Are measurable indicators defined?
- Is a time line or time frame for implementation indicated?

- Are human and financial resources indicated in the plan and are they available?

We have also developed materials on M&E for facilitators to use during the leadership program to support those responsible for providing M&E guidance to participating teams. Please see [http://erc.msh.org/leadernet/assets/files/M\\_E\\_Guide\\_to\\_VLDP\\_and\\_LDP\\_Final\\_2007.doc](http://erc.msh.org/leadernet/assets/files/M_E_Guide_to_VLDP_and_LDP_Final_2007.doc)

#### WHAT WE LEARNED FROM MEASURING WORK CLIMATE

MSH has used the WCA to measure work climate and climate change during most face-to-face leadership programs and all VLDPs conducted to date. Based on this experience, we offer several lessons learned about the use of the tool and the quality of the data it produces.

**Lesson 6: Improvements in climate cannot be attributed only to the effects of the LDP or VLDP due to intervening variables and the long time frame of the LDP.** Experience has shown that as teams learn to use leadership and management practices to address their challenges, work climate tends to improve. This is true for teams from NGOs and the central, provincial, and local levels of the public health sector. Nevertheless, there are limits to demonstrating changes in climate and attributing them to the intervention. The long time frame of the LDP (up to a year) means that other events, such as political or internal organizational changes, can impact climate for better or worse, despite the effects of the LDP.

In the ideal, our evaluations would use a control group; however, we cannot randomly assign participants to such a group. Participation in a leadership program is not indiscriminate and teams are not uninformed; teams that participate are either self-selected or are recommended by an organizational executive or administrator or a donor.

**Lesson 7: For programmatic and evaluation purposes, it is ideal to recruit intact teams for the LDP and VLDP.** The LDP tries to recruit intact teams. An intact team is one whose members have worked together before beginning the LDP and will continue to work together after the program ends (although all members of the team do not have to be in the same location). Intact teams are more likely to carry out their action plans, and the behaviors they adopt during the program tend to be sustained within the group afterwards. They are also more responsive to the WCA survey and easier to follow after the program ends.

Teams that are formed artificially for the purpose of participating in an LDP (for example, a group of supervisors from various levels of a public- or private-sector organization) have more difficulty in implementing their plans and measuring their progress after the program ends. These teams are often made up of individuals from geographically or administratively dispersed groups. They may work well together during the program, but when they return to their normal routines after its completion, they may no longer be motivated to work together as a team because they are not, in fact, a team.

Teams that disintegrate after the program ends are usually lost to follow-up, which negatively affects our ability to measure leadership outcomes. All members of a work group must fill out the WCA for evaluators to obtain a valid measure of work group climate. And because intact teams have a history of working together, they can respond to the WCA items in a meaningful way. Thus, not only for effective program design but also for evaluation purposes, it is preferable to recruit intact teams for leadership programs.

Despite the benefits of enrolling intact teams, new kinds of teams have emerged in the LDP and VLDP in response to the need to strengthen

multisectoral coordination. Examples include national- or district-level HIV & AIDS teams, teams working regionally to strengthen human resource management, teams newly formed by decentralization, and laboratory management teams.

If nonintact teams participate, the baseline WCA should be administered only after the teams have worked together for three or more weeks. For example, in working with new laboratory management teams in a program in Uganda, the LDP facilitators administered the WCA during the second LDP workshop instead of the first. In the VLDP the fixed structure of the online course makes such flexibility difficult. At the same time, the emphasis on using the baseline results of the WCA as an intervention in and of itself can be useful for new teams as they learn to work together.

**Lesson 8: Because the WCA is applied by program facilitators, the quality of the work climate data depends on their ability to explain to participants the purpose of the tool and how it is used.** Using work climate as an intermediate outcome of a leadership program can be challenging. Participants must understand how positive climate is created and that improved climate is an expected outcome linked to participation in the program. The purpose of using the WCA must be properly explained by program facilitators; otherwise respondents may not understand the value of the tool. Without this understanding, some participants may perceive climate as a measure that serves for reporting to a donor, not as an essential ingredient in developing an action plan and improving team performance. Moreover, if participants are not motivated to complete the WCA survey, the program evaluation will likely suffer from poor response rates.

For example, during the replication of the LDP in Egypt, when Egyptian facilitators delivered

the program to a new set of district and health facility teams, they did not use the WCA because they thought it was not essential to the program. As a result, the final evaluation lacked climate data to use for comparison purposes, although improved climate was one of the intended outcomes of the program.

To make the results of the WCA meaningful to teams, we have learned to weave the baseline results into the content of the leadership program. Facilitators help teams interpret their climate data and develop strategies to improve their climate using leadership practices.

Recently, the use of the WCA was greatly improved during a VLDP in Peru for staff from a municipality-based program, another USAID cooperating agency, and a local NGO collaborating with the program. The VLDP facilitator included a review of the WCA tool and analysis of the baseline climate results in one of the early modules and encouraged participants to discuss the tool, the importance of climate, and the actions needed to improve it. That VLDP also included pre- and post-course WCA surveys as a prerequisite for receiving the course diploma. Both strategies helped increase response rates substantially.

Finally, it is important that participants understand the five-point scale used in the WCA so they can respond correctly. This scale has occasionally caused confusion because participants are not familiar with scales or are not used to rating their experiences. Facilitators must carefully explain the scale and the meaning of the values. After the WCA, facilitators should help participants interpret their results and determine ways to improve their work climate. Finally, although MSH staff validated the WCA in different countries and languages, populations with low literacy and those not accustomed to self-administered questionnaires tend to have trouble responding. Some participants

may need extra attention during the application of the tool. For example, the leadership program in Mozambique relied on a local facilitator to apply the Portuguese version of the WCA among MOH personnel. This facilitator offered simultaneous translation into the local language for staff that were nonliterate or spoke a language other than Portuguese.

#### WHAT WE LEARNED FROM EVALUATING THE VLDP

M&L conducted follow-up evaluations of seven VLDP programs (three in Latin America, two in Africa, one in the Caribbean, and one in Brazil) to document the main outputs and outcomes produced by the participating teams. The evaluations focused on two elements: results achieved through the implementation of action plans and changes in work climate. Methods included a review of all action plans developed during the program and, where possible, an assessment of indicator data from the plans. In addition, we solicited responses to an e-mail questionnaire from the whole team as well as a representative from each team, and conducted telephone interviews with a member of selected teams. We also analyzed data from the application of WCA that were posted on the VLDP website by teams. Several lessons emerged from this experience about the most effective methods for obtaining information from virtual teams.

**Lesson 9: VLDP follow-up evaluations should take place six to nine months after program completion.** The timing of the follow-up evaluation is crucial for capturing high-quality information from participants. The more time that passes after completion of the VLDP, the lower the response rate to e-mail questionnaires and interview requests and the greater the risk of recall bias. At the same time, teams need enough time to practice new skills and implement their action plans before the follow-up evaluation takes place. The evalu-

ator must select a time frame that balances recall and results. In our experience, six to nine months after program completion is usually sufficient time for most teams to implement their plans.

#### **Lesson 10: VLDP e-mail questionnaires should be sent to a point person from the team, followed by a telephone interview with carefully chosen informants.**

The e-mail questionnaire solicits information from VLDP teams on measurable progress on the action plan (including indicator data) and processes the team used to implement its plan. Without adequate advance preparation by informants, however, the e-mail questionnaire is a poor mechanism for gathering process and outcome data. Our biggest challenge has been low response rates. We have tried different approaches to overcome this problem, with varying degrees of success. We have tried sending the questionnaire to the team leader alone; however, we found that this person often lacks information on indicator data. The information is much richer when all team members fill out the questionnaire, but response rates are usually poor when the questionnaire is sent to the entire team. The best approach we have found is to select a point person who will seek and compile responses from all team members. Teams need to be informed in the early stages of implementation of the program that e-mail questionnaires will be sent after the completion of the program. They should select a point person to disseminate the questionnaire and collect responses. But while using a point person has improved response rates, there is no guarantee that the questionnaire will actually be completed by all members of the team.

The next step is to conduct an in-depth telephone interview with one team member from each of several teams. The purpose of the interview is to verify the information from the

questionnaire and probe further about progress in results and work climate. The best strategy for selecting teams for telephone interviews is to categorize them into high, medium, and low performers according to criteria such as the quality of the action plan. Then two to three teams from each category are selected for the telephone interviews, often using geographic criteria. This process allows us to capture variations in the cohort and compare responses across performance levels.

Since the respondent's familiarity with the team's progress is essential for a high-quality telephone interview, each VLDP team should select a spokesperson who will be prepared to respond to telephone interviews or requests for information throughout the monitoring and follow-up evaluation period. The spokesperson should be enlisted early in the program and made aware of his or her responsibilities to represent the team.

## SECTION IV: Recommendations about Measuring Leadership

Effective leadership development is reflected by improved work climate, strengthened systems, and better performance in the delivery of health services. In addition, changes in behaviors, values, and norms take place in participating teams and their organizations, which are harder to make explicit. Our job as evaluators is to measure both the results achieved and explore the process used to reach them in order to develop well-founded insights into the likely causes of change and potential alternative explanations. We have learned that to do this well evaluators should keep several points in mind.

**Design the leadership program with the evaluation in mind.** Evaluations are highly dependent on the way in which the program facilitator introduces and uses measurement tools that supply data for the evaluation. For example, program facilitators who incorporate the WCA into the program in a way that is meaningful to participants can increase their motivation to respond to the questionnaire and use the results to monitor their progress in working together.

Participants need to be motivated to use the indicators in their action plans for monitoring performance and measuring change in relation to a baseline. Monitoring results needs to become a program strategy so that participants learn to value the collection and use of data to measure their progress. All leadership programs should use simple measurement instruments that participants will find useful in their work. This principle is especially important if the program relies on participant data to measure its successes or failures.

As we have noted, relying on the client organization's M&E system can either enhance

or weaken the ability to measure leadership outcomes. Participants from organizations with weak M&E systems are often unable to provide a suitable baseline measure. Or they may not use the indicators in their action plans to monitor performance, so post-program data are unavailable to measure change in relation to a baseline. This is particularly the case when the team has identified as its priority challenge an organizational process or system, which is usually assessed using qualitative methods. It happens less often when the team has identified a service delivery-related challenge, which can be assessed using available service data. The value of qualitative information has led us to strengthen our qualitative approaches and consider offering technical assistance in M&E to client organizations and participating teams.

**Balance the needs of donors without sacrificing the content of the evaluation.**

As program designers and evaluators, we must balance the need to report short-term results to donors and the need to orient participants to select significant targets that will contribute to their desired outcomes and can be measured after the program has ended. In general our leadership programs have been funded on an annual basis. This limitation has influenced data collection methods and the scope and timing of evaluations. Limited time frames often inhibit our ability to measure outcomes in terms of broad organizational change. Yet there is pressure from donors to report this kind of result during a short project. Furthermore, donors tend to prefer quantifiable evidence of outcomes even though qualitative methods are often more appropriate ways to capture change.

Programs can respond to these pressures in several ways. Programs should help participants develop performance benchmarks that can be

measured as the program is carried out, especially if outcomes are likely to change slowly and if it will be difficult to gather process data from participants after the program has ended. Alternatively, the immediate leadership outputs (leading practices) and intermediate outcomes (work climate and management systems) described above can be measured. Either approach may satisfy the need to report short-term data without undermining the teams' ambition to reach longer-term outcomes. Finally, more systematic reporting of results using mixed methods may meet donors' need for hard data, while highlighting the transformation process that occurred among participants and teams.

**Make use of open systems theory.** Causality is complex, especially when a result depends not only on improved leadership but also on other management capacity such as the availability of commodities or quality of supervision. There are multiple external influences on participants in a leadership program and results at each level (e.g., output, outcome) can be concurrent. This means that conceptual models that assume a linear relationship between inputs, processes, outputs and outcomes have a limited application. Although we have developed models to organize and plan our project designs and evaluation strategies, the fit of these models with leadership development is sometimes imperfect.

Taking an open systems approach may be the answer for the design of future MSH leadership projects and evaluations. We recognize that leadership development is an open system consisting of a complex interplay among

activities, individuals, teams, organizations, and communities. Our evaluations are already designed to take into account both the expected and unexpected results of this process. And our aim is not to attribute causality of results solely to program inputs but rather to measure contributions to results. In the future, we will consider using well-tested frameworks that take an open systems perspective, such as EvaluLEAD, as structures to guide the capture of qualitative information about and insights into the effects of leadership programs.

**Fully explore the leadership process to better understand the outcomes.** The difference between leadership development and other approaches to improving health services is in the process used. But to truly understand the leadership process, we must capture information about how people think, feel, and act.

For example, we have learned the importance of capturing information on what we call the "collaborative improvement effect." Through the collective analysis of their challenge and the selection and implementation of interventions to address it, team members learn from each other, develop new values and habits, and experience setbacks as well as progress in their selected health area. This collaborative improvement underscores the value of tracking intermediate outcomes and leadership outputs and explains why collecting participants' perceptions of the leadership process is vital to explaining outcomes. To evaluate a leadership program without capturing the voices and reflections of participants would be missing a key piece of the picture.

## Conclusion

Developing managers who lead in the health sector provides a fresh perspective on measuring performance improvement. One implication is that focusing more on the leadership process means producing larger amounts of qualitative information. In terms of analysis, it could be useful to determine whether the themes identified in the evaluation are reflected in existing models such as Kotter's Eight Step Model, which is an important underpinning of the content of the LDP (Kotter 1996, Kotter and Cohen 2002). An approach such as a fit-gap analysis could help frame the issues, guide the subsequent analysis, and better explain the evaluation findings.<sup>10</sup>

In our leadership programs, performance improvement occurs at two levels: intermediate changes within the team (work climate) and within the organization (systems) and outcomes at the service delivery level. The teams determine their main challenge and agree on their expected outcomes. They choose what is feasible to achieve and yet important enough to make a difference in the delivery of services. Then they set out on a journey that will revolutionize how they see each other as professionals, how they work together as colleagues, and how much they can achieve together. This paper describes how to capture the richness of the process and unearth some of the tangible accomplishments that result in order to inform our job as evaluators.

---

<sup>10</sup> Fit-gap analysis is a tool for carrying out a content analysis. This approach identifies parallels and discrepancies between the model of interest and the themes uncovered in the evaluation. Intersections between the model and evaluation themes can be mapped in a diagram or table to elucidate the findings.

## Resources

- Ceraso, Marion, et al. "Evaluating Public Health Leadership Training in Wisconsin: Beyond Participant Report." Abstract of a presentation (review of the literature on public health leadership training) at the American Public Health Association annual meeting, Nov. 2006. [http://apha.confex.com/apha/134am/techprogram/paper\\_136197.htm](http://apha.confex.com/apha/134am/techprogram/paper_136197.htm)
- Goleman, Daniel. "Leadership That Gets Results." *Harvard Business Review* March–April 2000, pp. 78–90.
- Grove, J., B. Kibel, and T. Haas. *EvaluLEAD: A Guide for Shaping and Evaluating Leadership Development Programs*. Oakland, CA: Sustainable Leadership Initiative, Public Health Institute, 2005. <http://www.phi.org/pdf-library/EvaluLEAD.pdf>
- Grove, J., and LEAG. "The EvaluLEAD Framework—Examining Success and Meaning: A Framework for Evaluating Leadership Development Interventions in Global Health." Oakland, CA: Population Leadership Program, Public Health Institute, 2002.
- Hannum, Kelly M., Jennifer W. Martineau, and Claire Reinelt, eds. *The Handbook of Leadership Development Evaluation*, San Francisco, CA: Jossey-Bass, 2007.
- Heifetz, Ronald A., and Donald L. Laurie. "The Work of Leadership." *Harvard Business Review* Jan.–Feb. 1997, pp. 123–34.
- Kotter, John P. *Leading Change*. Boston, MA: Harvard Business School Press, 1996.
- Kotter, John P., and Dan S. Cohen. *The Heart of Change: Real-Life Stories of How People Change Their Organizations*. Boston, MA: Harvard Business School Press, 2002.
- Laschinger, Heather K. Spence, Joan Finegan, and Judith Shamian. "The Impact of Workplace Empowerment, Organizational Trust on Staff Nurses' Work Satisfaction and Organizational Commitment." *Health Care Management Review* 2001, Vol. 26, pp. 7–23.
- LeMay, Nancy, Nancy Vollmer, and Alison Ellis. "Evaluating Leadership Development and Organizational Performance." In *The Handbook of Leadership Development Evaluation*, eds. Kelly M. Hannum, Jennifer W. Martineau, and Claire Reinelt. San Francisco, CA: Jossey-Bass, 2007, pp. 228–60.
- Litwin, George H., and Robert A. Stringer, Jr. *Motivation and Organizational Climate*. Boston, MA: Harvard University Press, 1968.
- Luoma, Marc, and Lauren Voltero. *Performance Improvement: Stages, Steps and Tools*. Chapel Hill, NC: IntraHealth, 2002. <http://www.prime2.org/sst/>
- Management Sciences for Health. "Creating a Work Climate That Motivates Staff and Improves Performance." *The Manager* 2002, Vol. 11, No. 3, pp. 1–22. <http://erc.msh.org/mainpage.cfm?file=2.1.55.htm&module=leadership&language=English>
- Management Sciences for Health. *From Vision to Action: A Course on Monitoring and Evaluation for Facilitators of Leadership Development Programs*. LeaderNet Workshop Resources, March 2007. [http://erc.msh.org/leadernet/assets/files/M\\_E\\_Guide\\_to\\_VLDP\\_and\\_LDP\\_Final\\_2007.doc](http://erc.msh.org/leadernet/assets/files/M_E_Guide_to_VLDP_and_LDP_Final_2007.doc)
- Management Sciences for Health. *Managers Who Lead: A Handbook for Improving Health Services*. Cambridge, MA: Management Sciences for Health, 2005a.
- Management Sciences for Health. *Work Climate Assessment Tool and WCA Guide for Facilitators*. 2005b. In the Health Manager's Toolkit at: <http://erc.msh.org/mainpage.cfm?file=96.9htm&module=toolkit&language=English>
- Perrin, Edward B., Jane S. Durch, and Susan M. Skillman, eds. *Health Performance Measurement in the Public Sector: Principles and Policies for Implementing an Information Network*. Washington, DC: National Academy Press, 1999.
- Perry, Cary, et al. "Validating a Work Group Climate Assessment Tool for Improving the Performance of Public Health Organizations." *Human Resources for Health* 2005, Vol. 3, No. 10, pp. 1–8. <http://www.human-resources-health.com/content/3/1/10>

Reinelt, Claire, et al. *Final Report: Leadership Matters: An Evaluation of Six Family Planning and Reproductive Health Leadership Programs*. Development Guild/DDI, 2005. [http://www.leadershiplearning.org/system/files/PackardGates\\_FinalReport\\_Web\\_0.pdf](http://www.leadershiplearning.org/system/files/PackardGates_FinalReport_Web_0.pdf)

Stringer, Robert A., Jr. *Leadership and Organizational Climate*. Upper Saddle River, NJ: Prentice Hall, 2002.

Umble, Karl, et al. "The National Public Health Leadership Institute: Evaluation of a Team-Based Approach to Developing Collaborative Public Health Leaders." *American Journal of Public Health* 2005, Vol. 95, No. 4, pp. 641–44. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1449235>

University of Southampton. *Leadership Evaluation: An Impact Evaluation of a Leadership Development Programme*. Southampton, UK: University of Southampton Health Care Innovation Unit and School of Management, 2004. [http://eprints.soton.ac.uk/14067/01/Leadership\\_Evaluation\\_report.pdf](http://eprints.soton.ac.uk/14067/01/Leadership_Evaluation_report.pdf)

## Acknowledgments

The authors would like to thank the participants in leadership development programs around the world whose contributions to evaluations formed the basis for the experiences and lessons documented in this paper. We would also like to thank our MSH colleagues for their vision and hard work in developing, refining, and conducting leadership development programs as well as for their commitment to evaluation. Thanks also to USAID for support of and funding for this Occasional Paper.

We welcome feedback on this paper. Please direct your comments to Nancy LeMay at [nlemay@msh.org](mailto:nlemay@msh.org) and Alison Ellis at [aellis@msh.org](mailto:aellis@msh.org).

All leadership development programs and evaluations described in this paper were carried out under the M&L and LMS programs with funding from the US Agency for International Development (USAID) Office of Population and Reproductive Health, Bureau for Global Health, award numbers HRN-A-00-00-00014-00 and GPO-A-00-05-00024-00. The opinions expressed herein are those of the authors and do not necessarily reflect the views of USAID.



**USAID**  
FROM THE AMERICAN PEOPLE

---

Management Sciences for Health  
784 Memorial Drive  
Cambridge, MA 02139-4613 USA

Telephone: 617.250.9500  
Fax: 617.250.9090  
Website: [www.msh.org](http://www.msh.org)