



learn fast. work smart. get results.

bp



**How do you measure the true value of a leadership development program? A global energy giant develops a model to find out.**

# BP Refines Leadership

Through a series of successful mergers and acquisitions, BP, the global energy company, had more than doubled in size in just four years. By 2000, the new BP was a more diverse and disparate organization, comprising employees from many different countries, nationalities, and corporate cultures. In name, BP was one entity; in reality, however, legacy companies within BP still conducted business in their own way. For the majority of BP's leaders, that inconsistency was obvious.

More than 35 different leadership development programs were taking place around the world, producing uneven results: Some leaders were well served; others were poorly served; and still others were not served at all. In addition, BP was missing an important opportunity: to develop a common understanding of what it meant to be a leader in BP.

## **The leadership challenge**

This condition presented a challenge for the core task force of three people, who among them brought more than 35 years of development experience to the table. The team decided to focus on the largest and most diverse population of leaders, now known as First Level Leaders. Totaling more than 10,000 people, more than two-thirds of the entire BP leadership population, these employees work in every business segment, including retail operations, chemical plants, refineries, and drilling platforms. They have diverse management responsibilities: Some lead teams of more than 10 people; others work with just a few subordinates in R&D, marketing, or human resources. They span a range of ethnic and cultural backgrounds, professional education, experience, and attitudes about work, the company, and life.

By Julie Brown,  
Ryan Eagar,  
and Paul Lawrence

Despite their differences, FLLs have one key factor in common: Their decisions, in aggregate, have an enormous effect on BP's turnover, costs, quality, safety, innovation, and environmental performance. In essence, this group represents a tremendous source of power and influence. Approximately 80,000 people report directly to them, including many employees who interact with customers.

The core team was given a mandate: to create an entirely new leadership development program for FLLs. Rather than locking themselves in a dark room, the team sought to include as many experts as possible. Among those invited to join was The Center for Strategy Research, a Boston-based research firm. CSR's in-depth, open-ended approach to research appealed to BP because it allowed for the full, unbiased, honest feedback that the team needed to better understand the challenges FLLs face. The team also wanted to partner with a firm where both parties could work together to construct this new leadership framework. According to one team member, "We liked that CSR didn't come in with a rigid methodology. And the team appreciated creating something together, which was different from what we each could have come up with individually."

Based on the findings of 2000, the First Level Leaders Development Program was launched in January 2002, built around the needs that FLLs expressed in worldwide workshops and telephone interviews. By the middle of 2002, nearly all of the local course offerings had been canceled.

## 2002: One language worldwide

The new FLL program was received with great excitement throughout BP. By the end of 2002, all targets set by BP's core team had been surpassed, and the program was hailed a success:

- Course attendees reported 84 percent satisfaction with the courses versus a target of 80 percent.
- Course demand was so high that more than 120 courses were offered, compared to the 100 courses that were originally planned. More than 4,500 FLLs attended a course, versus a target of 4,000 for the first year.
- The program was delivered at a unit cost more than 20 percent below plan.
- The organization experienced no resistance in engaging the most senior-level group leaders to host the events, despite their hectic agendas.
- Perhaps most significant, the BP team received a Helios Award, the highest internal level of recognition, from BP group chief executive, Lord Browne, who said, "The sharing of your knowledge and best practice across the group's different operations is recognized for transforming the way we do business."

## L&D team pushes for better measurements

Despite the program's success, the team was still uncomfortable. None of those measures addressed the program's effect on attendees over time. Furthermore, the value of the training couldn't be quantified or proven throughout the wider organization. The team's harshest realization: There had been neither a

# Remaining Relevant

"The richness of the data and of the insights that came out of the Center for Strategy Research's interviews has been greater than the team originally expected," says Cathy Dunn, FLL program lead, Organization Capability Team. "The process itself has been a useful one of re-engaging the population. In a sense, it is a form of intervention as well. It gives us a whole lot of data on how we can make the program more effective. We're convinced that these changes are helping the FLL program achieve a shelf life that is two to three times longer than the average training course within BP."

Some of the ways that BP has used the research to ensure FLL training remains relevant:

**Ensuring common experience.** Survey participants opined that Senior Level Leaders (FLLs' managers) must be trained on, or at least be familiar with, FLL training content and encourage FLLs to incorporate these lessons into their work. In response, Senior Level Leader training, which is currently in development, is incorporating several of the same principles and processes taught in FLL training.

**Ongoing refreshment and design of content.** In the initial rollout of the four-day leadership course, there were "mandatory" and "elective" modules. Survey participants shared their concerns that many of the issues that challenged them most were

covered in "elective" modules that ran concurrently, forcing them to choose between courses that would be equally helpful in handling day-to-day issues.

Many elective modules are now also being offered in training that is sponsored by other groups, such as the regions or the business segments themselves. That allows respondents multiple opportunities to attend the courses that hold the most meaning for them and ensures that, throughout BP, the same material is being taught, reinforcing "a common language."

**Reuniting to refresh attendee contacts.** In the surveys, participants shared their con-

rigorous discussion nor a consensus reached about the criteria for measuring the success of the program. Not having established a clear standard of measurement to assess the program's value, BP ran the risk that it would be difficult to prove the value of the training—not to mention the effect of divergent opinions within the company.

The team ultimately concluded that it needed to develop a way to prove the companywide impact and value of training, one that would take into account the diversity of experience and opinion within the company. With that goal in mind, the core team once again engaged CSR in January 2003 to brainstorm and address fundamental questions.

### Drilling deep into the challenge

One idea was to use the annual employee satisfaction survey to measure the effect of the program. That idea, however, was rejected because the employee satisfaction survey wouldn't give BP the ability to compare attendees with nonattendees, and it wouldn't permit BP to isolate the impact of the program from other events unrelated to the training.

BP and CSR reviewed the results and methodologies currently employed to measure the success of the FLL training, many of which are in use in other organizations. The team knew another challenge was ahead: changing ingrained corporate approaches toward assessing training programs. Until that point, the company rarely measured training in a systematic way, other than obtaining basic feedback informa-

Demand was so high that more than 120 courses were offered, compared



to the 100 courses that were originally planned. More than 4,500 FLLs attended a course, versus a target of 4,000 for the first year. The program was delivered at a



unit cost more than 20 percent below plan.

Photos: BP

cerns that, without some method to reconnect with contacts made during the courses, or some reminder of what they learned at FLL training, both contacts and content would grow stale.

Reunion events are now being sponsored at the region level to support connections made between individuals of different segments and to refresh key content.

#### Changing content to reflect changing needs.

More survey participants each year identify dealing with change as a challenge. In response, a module that had been offered as an elective course, Leading Change, is to be structured into the program as mandatory this year.

In addition, the importance of developing good management skills has grown since the initial survey was completed. In response, modules focusing on time management and delegation have been added to the Supervisory Essentials curriculum.

**Making content more applicable.** The initial rollout of the two-day Context & Connections course provided a detailed overview of BP's history and scope of operations. Participants found it interesting, but too broad to apply to their day-to-day jobs.

That feedback was shared with others within BP. Now, complementing FLL training, the business segments

are offering segment-specific guidelines on implementing the BP Management Framework and recommending tools that FLLs can use to ensure they are managing consistently with the expectations of their business lines.

**Improving morale.** For the past three years, one of the seven measurement criteria—the level of enthusiasm and excitement that an FLL brings to his or her job—has increased significantly among those who have been to training. That's in stark contrast to the results of an internal employee satisfaction survey in 2000 that expressed a disappointment in BP leaders.

tion such as how many participants attended a training session and what they thought of it.

But there was a sense that the “84 percent satisfied” rating didn’t tell the team what it really wanted to know: not only what the First Level Leaders recalled from the training experience, but also what they actually put into practice over an extended period of time.

Obtaining that kind of information required a survey effort that takes place well after the fact, and regularly thereafter. The team concluded that the “happy sheets,” the short satisfaction surveys completed at the conclusion of a course, wouldn’t help to address the real research question: When you got back to the office after your training, what happened?

If and when the response was “nothing,” the team wanted to understand what managerial support would be required and what barriers would need to be overcome to enable FLLs to implement what they learned.

There was a sense that the “84 percent satisfied” rating didn’t tell the team what it really wanted to know.

The team also wanted to better understand how training experiences were applied to workday situations. One goal was to gather from attendees, their managers, and their reports, examples of instances in which training was applied and resulted in improved outcomes.

Finally, a critical goal of the research project was to re-engage participants in the development process. CSR used a research methodology that reminded participants of their training experience and actively engaged them in a discussion about it during the interviews. The team believed that an ongoing program of contacting and monitoring the graduates of the training would help address the physical and emotional isolation some of those leaders experienced while working in remote areas of the world.

### A new level of evaluation

While applying Donald Kirkpatrick’s classic model of evaluation, the task force discovered that the organization was caught somewhere between sub-Level 1: Reaction and Level 1: Reaction. (Do the participants like the program?) The former, where BP often operated, suggests that in some cases the evaluation didn’t

get as far as Level 1 because measurement criteria consisted primarily of number of programs run per year and number of attendees per program.

Applying Level 2 of Kirkpatrick’s model (Did the participants learn?) became more difficult, especially in the area of leadership. Unlike other areas, such as legal, compliance, or licensable skills, asking leaders to attend a leadership course and then pass an exam or obtain a license in content is a challenge. By its nature, the subject matter of leadership isn’t well suited to that straightforward manner of testing the learner to validate knowledge gained.

To this point, feedback about any leadership training that went beyond satisfaction measures was anecdotal. While the team heard how margin-generating initiatives had begun in the leadership development courses, senior leaders often interpreted that information with skepticism. There was no way of proving what would have happened otherwise.

Looking at the next level in Kirkpatrick’s model, Level 3 (Are participants using what they learned?), the team felt that approach was worth building, as long as it had a focus on business benefits. Unless an aspect was added that proved those benefits, the perception was that some organizational leaders would regard stories about behavioral change as unrelated to the hard-edged conversation about bottom-line performance.

The team feared trying to establish statistical correlations between business performance and program attendance. Even if statistical models could be built to reliably show correlations, the team knew that any claims based on those results would again be treated with skepticism. So, the team rejected a straightforward Level 4: Results approach. (Does it have bottom line impact?)

Although the team was intrigued by Levels 2, 3, and 4, the members wanted to adopt an approach that wasn’t perceived as coercing the trainers into delivering bottom-line dollars. That perception would be at odds with the notion of leadership as a matter of individual choice.

### The WIN model

The team devised a two-stage process to answer the critical question: How do we establish a link between the annual training investment of almost US\$4 million and bottom-line performance?

The goal wasn’t to establish an objective “truth” that would withstand the challenge of a disinterested academic. Instead, the team needed to develop measures that the organization believed in. If the team could show that the program was having an effect on specific behaviors—that is, the behaviors that the organization believed would have a direct impact on bot-

tom-line performance—that would be the most powerful argument of all. To that end, the team devised its own approach and called it the WIN model.

### Stage I: Identifying expectations

The objective of this stage of WIN is to identify, through open-ended inquiry, the key behaviors that those most affected by the training would expect to see changed by a successful leadership development program. CSR and BP worked together to develop an in-depth qualitative telephone survey. Translated in four different languages, the survey was completed by 300 employees worldwide. Based on the findings, seven categories of behavioral changes were identified.

The importance that each level of leadership throughout BP attached to each of those seven categories was also quantified. Group leaders, for example, rated the importance of team performance and organizational awareness higher than did other populations. Senior leaders, the FLLs' managers, rated management skills, confidence, and self-awareness relatively high in

importance. FLLs themselves and their direct reports rated communication skills highest in importance.

Those varying levels of importance resulted in a weighting strategy that would help prove the effect of training through the organization.

### Stage II: Measuring against expectations

Once the key measures had been established and ranked by different levels of the organization, BP implemented the second stage of the WIN model. Using a combination of 300 telephone and 1,200 Internet surveys, employees were asked to evaluate FLLs according to those established measures. FLLs were asked whether they had attended a training event and asked to evaluate themselves on each measure. Direct reports and FLLs' managers were asked if their FLL had been to training and then asked to evaluate him or her. Two separate populations were established: FLLs who had been to training and those who had not. The results of the two groups were compared.

### Surprising results

## First Level Leaders

### 2000

**67 Different Programs Worldwide** covering 1,670 attendees annually at an average cost of US\$2,500 per attendee

#### Research Results of FLLs

- "BP is doing so much now that it's very difficult to remain focused on particular areas, and understand what is most important, and why."
- "That concerns the members of my team most, not knowing what the future holds, not knowing what the rules of the game are, because the rules of the game keep changing."

#### Research Results of FLLs

- "In the last four years, I never had professional training, never even had meetings or teleconferencing. By sharing our best practices or information, I think I can develop new ideas, new ways of doing things. Many things I deal with, I know someone at BP has dealt with, but I feel quite alone on this."
- "How to manage people. The team management side of it. What are some of the things you could arm your First Line Leaders with that would help maximize their performance and get the most out of their teams? That I just saw as lacking."

#### Research Results of Direct Reports

- "We have capabilities you may be missing. You can delegate more to us. We are responsible."

### 2002

**3 Different Programs Worldwide** covering 4,500 attendees annually at an average cost of \$880 per attendee

#### Context & Connections

- 2 1/2-day course:
- Organization Structure
  - Performance Goals
  - BP's Global and Regional Strategy
  - The Brand and What BP Stands For

#### Leadership Event

- 4-day course:
- The Role of Leadership in BP
  - The Role of the FLL
  - Leading Self, Leading Others
  - Leading Teams

#### Supervisory Essentials

- E-learning modules and locally sponsored programs:
- Understanding Financial Decisions
  - Safety Leadership
  - Local HSSE (Health, Safety, Security, Environment) Leadership and Legislation
  - Employment Legislation
  - Ethics

# Behavioral Changes

BP identified seven categories of behavioral changes. Here are their descriptions, as defined by survey participants.

## Organizational Awareness

Understanding the wider BP organization and its goals; the ability to network with peers across the company.

## Interpersonal Skills

Ability to work well with others and to recognize and act upon the developmental needs and career aspirations of reports. Awareness of others' feelings, needs, and concerns.

## Communication Skills

The abilities to listen well, understand, and take into account others' viewpoints; to compose and convey ideas persuasively; and to communicate with others in the organization, including peers, reports, and management.

## Confidence and Self-Awareness

The ability to take an objective look at the way one does things in order to identify areas to improve; to have confidence with respect to decision making; to recognize how personal style impacts effectiveness; and to remain calm in the face of challenges.

## Management Skills

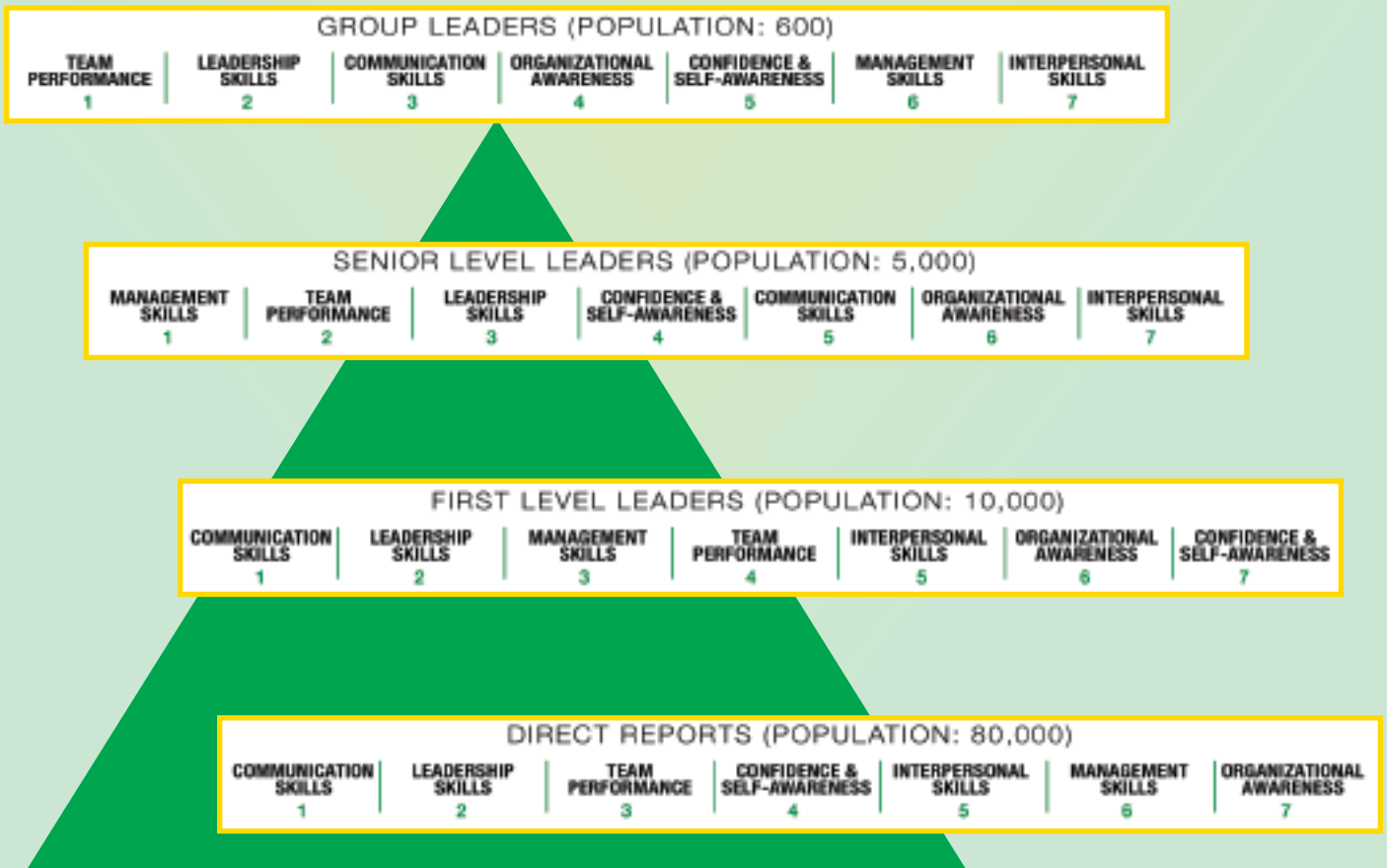
Abilities in problem solving, prioritizing, delegation, and managing one's own time effectively.

## Leadership Skills

The abilities to develop, maintain, and deploy an effective team; diagnose shortcomings associated with the team; creatively overcome barriers and obstacles to meeting the requirements of a job or project; and generate commitment to common goals.

## Team Performance

The ability to lead the team to achieve outstanding results.



Both the FLLs' direct reports and their managers ascribed statistically significant better performance on every measure for those First Level Leaders who had attended the training events, compared to those who had not.

Most surprising, while FLLs themselves did report some improvement on all measures, that improvement was significant on only three of the seven measures. The most significant differences were noted by FLLs' direct reports, who detailed scores that were 14 percent higher for trained FLLs.

In addition, the conversational telephone interviews provided context, depth, and success stories — examples of FLLs putting training to good use. That provided the team with useful anecdotes for conversations with leadership, internal publications, and other uses.

Instead of FLLs reporting improvements in their own competencies, with the rest of the organization perhaps not fully aware of any change in behavior, the opposite was found. Those closest to FLLs noticed improvements in behavior, while FLLs themselves reported it less clearly. The implication is that although the program has a significant effect on First Level Leaders, they are unlikely to reveal the entire story. It's critical to survey those people who come in direct contact with the FLLs to get a full understanding of the effects of training on that population.

### **WIN model stands test of time**

A year later, in April 2004, BP and CSR implemented the third stage of the WIN model, an annual measurement and benchmarking survey. Again using a combination of telephone and Internet surveys, FLLs were asked whether they had attended an FLL event and to evaluate themselves on each measure.

Direct reports and managers were asked whether their FLL had been to an event and to answer the same questions about him or her. Again, the performance ratings of those who had been to training were compared to those who had not.

These results show a similarly compelling story to the previous year—as well as a key difference. The direct reports again ascribed significantly better performance on every measure for FLLs who attended training. Given the vast number of people who report to FLLs, measuring such improvements in management quality suggests the FLL events are having considerable impact. And again, FLLs' managers reported significantly better performance on most measures.

In 2004, however, FLLs reported statistically significant better performance, on five of the seven measures evaluated.

Yet, that difference was still far narrower than the

differences reported by subordinates and managers. Again we see that surveying training attendees alone is unlikely to reveal the full effect of training. It appears to take more time for those who actually experienced the training to be able to evaluate the ways training has been useful.

### **The next step**

As more First Level Leaders attend at least one of the FLL training courses, the control group of untrained employees will become smaller and less statistically meaningful.

In 2005, the model will compare the results of those who have attended no courses to those who have attended one or two and three or more courses. Interestingly, the 2004 research also showed that those who had attended three courses scored significantly higher than those who attended one or two courses.

Proving the value of training to an organization is not easy. In the case of the FLL program, it seems prudent to spend an incremental small percent of its annual \$4 million investment to ensure that the training continues to meet the ongoing needs of the organization and that those results can be measured and proven.

In addition, given the vast number of FLLs and the even more wide-reaching number of people they influence (their reports, customers, vendors, and strategic partners), the measured progress that those managers are making through training has an impact through a large number of people affiliated with BP.

In retrospect, the WIN model represents a logical extension of the approach that had already underpinned the success of FLL. Instead of basing activities, and conclusions about the impact of those activities, on theoretical constructs and models, the WIN model was based on establishing tangible, well-structured, and meaningful conversations across the organization. The result: an evaluative method that BP believes in. That method is based on an ongoing and dynamic approach, rather than any statistical methodology that relies on historic correlations to make assertions about the future. **TD**

**Julie Brown** is president of The Center For Strategy Research, Boston; [jb@csr-bos.com](mailto:jb@csr-bos.com). **Ryan Eagar** is a leadership advisor, organizational capability, BP International Ltd., Sunbury-on-Thames, Middlesex, England; [eagarr@bp.com](mailto:eagarr@bp.com).

**Paul Lawrence**, formerly of BP, is a learning and development consultant based in Sydney, Australia; [paul\\_lawrence22@yahoo.com.au](mailto:paul_lawrence22@yahoo.com.au).



Reprinted with permission from T+D Magazine, March 2005, by The Reprint Outsource, 717-394-7350

Copyright 2005 by The American Society for Training & Development  
1640 King Street, P.O. Box 1443, Alexandria, VA 22313-2043, Phone 703.683.8100. All rights reserved.