A Practical Guide to Evaluating Coaching: Translating State-of-the-Art Techniques to the Real World

David B. Peterson

Kurt Kraiger

Our objective for this chapter is to provide a practical framework for examining the effectiveness and the value of organizational coaching programs. The principles we present apply to any type of coaching, although we primarily address one-on-one coaching provided by formally designated coaches charged with helping people gain insight, learn new skills, and improve on-the-job performance.

In an ideal world, the evaluation process is integrally connected to the implementation process, and both are driven by clear purpose and design. Reality, though, is quite a bit different. Kraiger (2002) summarized key factors that diminish the likelihood that any training and development program is evaluated. These factors, which also apply to the evaluation of coaching programs, include the additional costs associated with conducting the evaluation, the evaluator's lack of expertise in statistics or research methods, and a lack of time. In addition to these

constraints, coaching may be even more difficult to evaluate because it is individually customized, shrouded in issues of confidentiality, delivered by a range of coaches who rarely follow standardized procedures, and frequently implemented piecemeal rather than systematically throughout the organization. Because of this last point, the typical evaluation process for coaching arises after the fact, when someone observes, "Hey, we're spending a lot of money on this. Let's find out what we're getting for that investment."

Further, the professional practice of coaching is still maturing. Leaders in the field are debating virtually every aspect of coaching, including the very definition itself (Kampa & White, 2002; Peterson & Hicks, 1999; West & Milan, 2001). This lack of consistency makes it essential that users of coaching understand how to evaluate the services they are receiving. This chapter therefore presents a hands-on, practical process for measuring the impact of coaching on the individual participants and on the organization. We begin with an overview of the research that has been conducted on coaching, move to a discussion of the major challenges and issues in evaluating coaching, and conclude with a step-by-step blueprint for designing your own evaluation process. Even though we present a systematic approach to the process, we understand the constraints that operate in most environments and so provide simple and quick options to suit a wide range of needs.

Research on Coaching

Coaching works. That much is simple. Based on dozens of case studies, hundreds of personal testimonials in scores of organizations, and diverse threads of research (Birkeland, Davis, Goff, Campbell, & Duke, 1997; Davis & Petchenik, 1998; Gegner, 1997; Kampa-Kokesch, 2001; McGovern et al., 2001; Peterson, 1993a, 1993b; Thompson, 1986; see also Bloom, 1984), it is clear that coaching has an impact on people and on business results. However, anyone who has worked in human resources (HR) for very long understands that answering the question of whether an intervention works by stating that it has been proven effective in other organizations may not be a sufficient answer. There are two reasons for this. The first is that executives often believe that the effectiveness of certain techniques varies from organization to organization. Therefore, you will want to read the rest of this chapter to learn how to design an evaluation that fits the coaching program in *your* organization. The second reason is that executives evaluate an intervention more broadly than just by its effectiveness or whether it worked (Boudreau & Ramstad, 2003). For many of them to be persuaded that HR programs support business goals, they also require evidence of the impact (did it make a strategic difference?) and efficiency (was it worth the cost?) of the intervention (Boudreau & Ramstad, 2003). Therefore, you will also want to read the rest of the chapter to learn how to conduct evaluations that address the effectiveness, impact, and efficiency of coaching programs.

For now, when we say that coaching works, we are commenting on its *effectiveness* as a method for helping people improve their performance in areas of leadership,

communication, interpersonal, and cognitive skills (Peterson, 1993b). Similar to return on investment (ROI), impact and efficiency can only be measured in context, since they depend on the strategic value of improving a specific capability (impact) and the value of a specific person's time (efficiency). This point is discussed in more detail below.

With that in mind, consider the coaching research that has been done. Thompson (1986) is the earliest known study examining behavior change in actual coaching participants. His chief findings, based on 150 managers and executives from a range of organizations, were that participants' bosses reported significant behavior changes due to coaching and that those changes lasted for at least 1 to 2 years following the coaching engagement.

Peterson (1993a, 1993b) studied 370 coaching participants, finding an average effect size of 1.56 or greater on specific learning objectives as rated by participant, manager, and coach. That is, coaching produced an average change of at least 1.56 standard deviation units, roughly the equivalent of moving from the 50th percentile to the 93rd percentile of performance. These results compare favorably to the average effect size of approximately 0.45 reported by Burke and Day (1986) in their meta-analysis of management training and development programs. Based on this comparison, a case can be made that coaching is at least three times more effective than the typical training program in producing results.[1] In addition to measuring improvement on specific coaching objectives (i.e., effectiveness), Peterson obtained measures of overall job performance (i.e., impact). Both boss and participant ratings showed an increase in effect size of .08 or greater for overall performance, indicating substantial incremental value to the organization. These gains were still evident 1 to 2 years after the completion of coaching (Peterson, 1993b).

In 1997, AMOCO Corporation (now part of British Petroleum (BP) conducted a historical evaluation of the coaching they had offered to their executives over the preceding 10 years (Birkeland et al., 1997; Davis & Petchenik, 1998). Noting the problematic nature of this type of research (e.g., poor performers and participants with negative coaching outcomes may have left the organization by the time of the follow-up survey), the authors concluded that "caution should be taken before placing too much stock in any single piece of evidence that was presented. However, throughout the study, the results *consistently* suggested that coaching facilitates the development of skills which positively affects both the individual participant and the overall organization. It is this consistency in results which enables us to reach the conclusion that the individual coaching program has been a valuable investment for AMOCO" (Birkeland et al., 1997, p. 15).

Compared to other AMOCO managers, coaching participants consistently demonstrated improved performance, increased ratings of potential for advancement and 50% higher average salary increases. Also important, participants themselves attributed these results directly to the coaching they had received.

McGovern et al. (2001) conducted one of the first studies to look exclusively at the impact of coaching, examining 100 mid- to senior-level managers at large organizations who participated in coaching. They estimated their coaching to be worth 5.7 times the initial investment, based on a conservative formula for estimating ROI

developed by Phillips (1997). Organizational benefits from coaching included better productivity (reported by 53% of participants), quality (48%), organizational strength (48%), customer service (39%), and retention (32%). They also cited improvements in cost reduction (23%) and profitability (22%). Among personal benefits, they reported improved relationships with direct reports (77%) and peers (63%), better teamwork (67%), and increased job satisfaction (61%).

In conclusion, these and other studies (summarized in Kampa & White, 2002; Kampa-Kokesch & Anderson, 2001) provide substantial evidence that well-designed coaching makes a meaningful difference in skill learning, job performance, and organizational results. Nonetheless, we advise caution in generalizing these results—the term *coaching* is used to represent an extensive range of services, and it is doubtful that all providers of coaching will yield comparable results (Kampa & White, 2002; Peterson & Hicks, 1999). In fact, there are examples showing that coaching has not been an effective intervention. One theme in these examples is utilizing a coach who relies primarily on just one methodology for coaching (e.g., 360-degree feedback or other insight-oriented approaches) and does not guide participants through the full development process (see Peterson, 2002, for a more detailed treatment of the necessary conditions for learning). Even when the coaching is well designed and participants are carefully screened, coaching may fall short. Although the average change reported in Peterson (1993b) is substantial, two of the 370 participants did not show any measurable improvement. A final concern is that few studies have connected coaching outcomes to objective measures; virtually all research has been based on self-report and observer ratings. Therefore, it is essential for organizations to conduct some type of evaluation to ensure that coaching is meeting their needs.

Challenges and Issues in Evaluating Coaching

Although research evidence shows that coaching works, these studies also illustrate many of the challenges of conducting quality evaluations of coaching. In this section, we highlight three critical issues in evaluating coaching; in the final section, we lay out a step-by-step process for designing evaluations. In addressing these issues, we draw on existing theory and methods of evaluation in training and organizational development (e.g., Kraiger, 2002; Martineau & Preskill, 2002), previous evaluation research on coaching, and our own experiences working with a variety of organizations.

Purpose

The first challenge in evaluating coaching is determining the purpose: What questions need to be answered and what decisions need to be made? Because evaluation may consume financial resources and participants' time, it should be undertaken with a clear understanding of what is to be learned from the evaluation. Yet more often than not, it is our experience that organizations do not clearly think

through the consequences or benefits of evaluation before initiating such efforts. It is said that the unaimed arrow never misses its mark, but in modern organizations, that archer will not get a second chance to shoot.

Kraiger (2002) suggested three primary purposes for evaluation:

1. *Decision Making:* Using evaluation outcomes to change the status of a coaching program, its participants, or its coaches (e.g., continue, expand, or scale back a program, or reassign a coach)
2. *Improvement:* Providing evaluation outcomes to coaches or participants in order to have a positive impact on their participation or to improve the process they are following (e.g., indicate the need for more frequent meetings)
3. *Marketing:* Using evaluation outcomes to encourage adoption of a coaching program by a new organization, garner senior management support, or encourage potential participants to enroll

Deciding on the purpose for evaluation will help to ensure that the right information is collected. For example, if the purpose is to evaluate the quality of the coaches, you will formulate different questions depending on whether you are interested in making decisions about who to use, sharing feedback to improve the quality of their work, or marketing their results to others. For obvious reasons, it is important to consider the needs and preferences of the stakeholders, such as executive decision makers, sponsors of the coaching program, coaches, or potential coaching participants, who will use the data that are collected (Martineau & Preskill, 2002).

Design

The second issue is choosing an appropriate research design. In our experience, even though many of the managers and executives who see evaluation results lack formal research and statistical skills, they are adept at generating alternative explanations for observed effects (e.g., speculating that increased productivity could be due not to coaching but to seasonal variations or staffing changes). Therefore, one of the best strategies for improving the quality of coaching evaluations is to develop a sound research design that minimizes the possibility of alternative explanations.

Thoughtful design makes positive outcomes more compelling. Suppose that a group of 10 newly promoted executives all received coaching for 9 months. At the end of this period, these executives reported an average improvement of 50% in their "sensitivity to the needs of others" and a 40% increase in their "capacity to see the big picture" when making decisions. Although this evidence appears to reflect positively on the coaching program, an obvious counter argument can be made—how do we know that these executives did not develop naturally (on their own) without any benefits from coaching, perhaps because the promotion itself stimulated their development?

Had the organization included a control group—a matched set of executives who were promoted but did not attend training—it would be in a better position

to isolate the benefits of coaching. Sometimes such control groups may occur naturally in an organization. However, in other situations, withholding treatment can be unreasonable for political and ethical reasons.

The use of pretest or baseline data is another effective strategy for establishing a context for performance improvement. For example, if the subordinates of our 10 executives rated the interpersonal skills of their bosses as "very effective," this evidence is more persuasive if we know that the same subordinates rated the same bosses as "ineffective" just weeks before the coaching began. However, pretest data may not be available in all evaluations, such as when the need to evaluate occurs *after* the coaching has begun.

When the use of a control group and/or pretest data is not possible, an effective alternative is to use evaluation items as a control (Haccoun & Hamtiaux, 1994). Peterson (1993b) provides an example of this strategy. Two sets of items are used to evaluate the outcome of coaching: Items that measure skills targeted by coaching (e.g., interpersonal skills) are compared with items that measure managerial skills not targeted by coaching (e.g., financial management skills). Both sets of items can be evaluated by ratings or objective performance indices if available. Data showing that coaching-related items improved significantly more than non-coaching items would be strong evidence for the effectiveness of coaching (Haccoun & Hamtiaux, 1994; Peterson, 1993a). In Peterson's (1993b) study, control items in fact showed no significant change when measured at the completion of coaching or at follow-up 1 to 2 years later. This technique can be used even after coaching has been completed, or whenever only post-coaching rating data can be collected. By rating participants on the amount of improvement they have demonstrated, using both sets of items, you can calculate the effect due to coaching.

Thoughtful design consists of anticipating alternative explanations for outcomes and providing a means for rationally arguing whether coaching was the most plausible explanation for the observed effects. As McLinden (1995) noted, organizational decision makers do not expect research to *prove* that a program works but to provide reasonable evidence that the program *could be responsible.* Over time, following a well-designed evaluation process, consistent evidence of improvement from a variety of participants under a variety of conditions can build solid support for the coaching program.

Return on Investment and the Impact of Coaching

This leads into the third challenge: how to show whether coaching makes a difference. Questions about ROI are frequently asked by HR professionals interested in proving to line managers that investments in coaching are worthwhile. Even though it is tempting to cite McGovern et al. (2001) and give a single numerical answer to the question, it is essential to realize that ROI (and any measure of impact) is context specific. For example, using coaching to help leaders in two different organizations improve their ability to think strategically may cost the same (efficiency) and produce the same level of learning (effectiveness) and yet have negligible bottom line impact

in one organization and huge impact in the other. These seemingly conflicting results occur simply because the relative importance of strategic thinking to each organization's success is different. For that reason, it is impossible to derive a universally applicable number for "the ROI of coaching."

Although there are methods for estimating the economic value of new behaviors and relating these to program costs (see Cascio, 1982; Phillips, 1997), the most compelling support for coaching comes from analysis of the business impact, effectiveness, and efficiency of the program (Boudreau & Ramstad, 2003).

To understand the potential business impact, you might ask questions such as these: Where does lack of talent limit our ability to achieve strategic business objectives? What skills are most critical to business success, and which of these are in shortest supply? Where would improvement in the organization's talent have the greatest impact on success? It is also crucial to ask about the relative value of the impact, such as, How much would it be worth to achieve that improvement in talent? For example, say the critical issue for senior leadership is the lack of general managers who can think strategically and provide decisive leadership in times of constant change. Senior executives tell you that an effective leader in that role is worth twice what a mediocre leader is worth. If the typical salary in that role is $100,000, it is not unreasonable to contemplate a $50,000 per person investment for the intervention. Getting executives to appraise the value of a performance improvement is one of the keys to determining ROI.

Second, to examine effectiveness, consider the range of solutions (including coaching) that are available for each skill/performance area identified. How effective is each possible solution in generating improvement for that skill? As a broad generality, it is reasonable to say that to produce measurable improvements in performance, coaching is more effective than classroom training, which is more effective than reading a book. Of course, coaching can easily be more expensive than group training, which in turn is generally more expensive than reading a book. Keep in mind that ROI is a function of inputs (investment) as well as outcomes (return), so all costs need to be taken into account, including administrative support and the participant's time. In many instances, the opportunity cost of an executive's time is higher than the actual dollar cost of the book they are reading, or even the training or coaching program they are participating in. In our example, the key question is whether any of these solutions is capable of taking a general manager with mediocre strategic and leadership skills and turning him or her into a strong leader.

Finally, to examine efficiency and arrive at a determination of ROI, consider how to allocate resources optimally to accomplish key objectives. This is a function of the amount of change that is necessary and the financial and nonfinancial cost of the intervention. If only small changes are necessary, less costly and less powerful solutions (such as 360-degree feedback, on-the-job experience, peer coaching, books, or Web-based resources) might be sufficient. When dramatic, rapid change is essential, a powerful solution such as coaching may be the only satisfactory option. At this point, it is appropriate to choose the cheapest solution that can generate the desired amount of change. In our example, coaching could be recommended as long as the

total financial and nonfinancial costs (including administration and the manager's time) are under $50,000 per person. The low-cost option of reading a book is unlikely to transform managers into strategic leaders, so even though it is cheap, the ROI in this hypothetical example is virtually zero.

The real question is not what is the ROI of coaching, but what solution is sufficiently effective at the lowest cost to accomplish the objective, whether that is coaching, a book, or a Web-based program. Too many conversations start from the wrong direction, with questions such as these: How do we justify the amount of money we are spending on coaching? Are we getting our money's worth? Searching for ROI is often an attempt to justify a decision that's already been made. Advocates of particular solutions, such as coaching, lose credibility with line executives when they cannot provide the necessary evidence of the program's value. In contrast, credibility comes from starting with the needs of the business and finding useful solutions. Given the type of capabilities we need for our business, what solution produces the change we need at a reasonable cost? This puts the onus on business leaders to determine the value of the need and places HR in the critical role of providing cost-effective solutions.

A Practical Guide to Evaluating Coaching

The previous section of this chapter discusses some of the challenges and issues in planning and conducting an evaluation of coaching programs. This section presents a practical, hands-on approach to evaluation. The key points in this section are summarized in Figure 13.1.

Step 1: Lay the Foundation

As can be seen in Figure 13.1, there are five steps to planning and conducting a successful evaluation of any coaching program. Step 1 sets the stage for an effective evaluation by thinking through the purposes, resources, and challenges of evaluation. Ideally, the evaluation process is designed in tandem with the design of the coaching process. However, the guide presented here can be followed regardless of whether evaluation is sought before, during, or after the coaching has taken place.

First, decide on the purpose for conducting the evaluation. Consider not only the immediate needs, but future needs as well. Well-done evaluations are relatively rare, so consider whether there is a way to leverage the evaluation to meet secondary or future needs related to decision making, marketing, and improvement. For example, the primary purpose of the evaluation may be to provide cost-effectiveness data in the event of anticipated budget cuts, but consider how additional data collected during the evaluation may provide useful feedback to coaches or participants. Explain the purpose to everyone you talk with while planning and conducting these evaluations to refine your focus and to obtain their buy-in. We are aware that HR is occasionally given a mandate by senior management to evaluate a

Evaluation Process Checklist

Step 1: Lay the foundation

- ❑ Decide on the purpose for evaluation
  - – Improvement
  - – Decision making
  - – Marketing

- ❑ Identify the audience for evaluation results
  - – Coaches
  - – Potential/future coaches
  - – Participants
  - – Potential/future participants
  - – Stakeholders and decision makers

- ❑ Marshal resources for evaluation
  - – Financial resources
  - – Personnel resources
  - – Time resources

- ❑ Anticipate and plan for obstacles
  - – Resource-based
  - – Motivational
  - – Design-based

- ❑ Anticipate and plan for cultural, diversity, and language issues

Step 2: Design the process

- ❑ Choose raters
  - – Coaches, participants, bosses, subordinates, peers, customers
  - – Other stakeholders

- ❑ Design content for evaluation
  - – Coaching objectives for each individual (individually customized and/or common items)
  - – Targeted organizational objectives for coaching
  - – Aspects of the coaching process and the coach
  - – Suggestions for improving the process

- ❑ Choose response scales
  - – Yes/no
  - – Agree/disagree
  - – Current effectiveness
  - – Level of improvement
  - – Current vs. desired performance

- ❑ Determine data collection process
- ❑ Design evaluation
  - – Pretest/posttest options
  - – Control/comparison group
  - – Item-level controls

Figure 13.1   Evaluation Process Checklist

Step 3: Implement the process

- ❑ Distribute items to be rated/answered
- ❑ Collect ratings/responses
  - – Follow up as necessary to improve response rate
- ❑ Store ratings/responses
- ❑ Maintain confidentiality and integrity of the data

Step 4: Analyze the data

- ❑ Account for attrition
  - – Determine dropout rates
  - – Determine differences between participants who dropped out and participants who completed
- ❑ Evaluate impact
  - – Is the program consistent with strategic objectives?
  - – Do coaching objectives support these objectives?
- ❑ Evaluate effectiveness
  - – Were program objectives met?
  - – Have participants improved?
  - – How effective are the coaches?
- ❑ Evaluate efficiency
  - – Are benefits worth the investment in coaching?
  - – What is the total cost?
  - – Can program costs be scaled back?

Step 5: Present the findings

- ❑ Link results back to purposes for evaluation
  - – Identify specific purposes for the evaluation
  - – Discuss results as answers to specific questions
- ❑ Executive summary
- ❑ Graphical vs. statistical data
  - – Present results in different levels of detail
  - – Tailor presentation to preferences of audience

Figure 13.1   Evaluation Process Checklist

program; it is even possible that such a directive may arrive without a clearly articulated purpose or goal. In such cases, you will often have to play a consultative role in helping senior management clarify exactly what they are trying to accomplish. Be assured that your efforts will be rewarded when you present the results of the evaluation.

Two general questions prompt many of the evaluation efforts around coaching:

1. Does coaching really work? Here, the real question is almost always, Will coaching work in our organization, with our people, to give us the results that we need?
2. Are we getting our money's worth? What's the ROI of coaching in our organization?

Other reasons to conduct evaluation include improving the organization's coaching process (examining factors such as coach and participant selection, coach-participant matching, organizational and managerial support, environmental and organizational obstacles), determining the best pool of coaches for the organization's needs (areas of expertise, knowledge of the organization, cultural fit, perceived credibility by potential participants, etc.), and improving the success rate of those selected to participate in coaching (personal motivation, purpose, organizational support, etc.). For best practices and other important aspects to consider in designing and evaluating a coaching program, see Executive Coaching Forum (2001) and Peterson (2002).

Second, determine the various audiences for the results of the evaluation. Consider who will see the results of the intervention and the type of evidence they require. Will coaches and participants want to know how well they are doing or how effective the program is? Are there champions for coaching within the organization or sponsors for the evaluation? What other stakeholders or decision makers will want to know about the evaluation results? How can evidence of the success of the program be used to enhance the participation of future coaches or participants? As you make decisions about the intended audiences for the evaluation, you will want to understand what they hope to learn from the evaluation and involve them in the evaluation planning process. A belief persists that top executives are primarily influenced by evidence of financial impact such as shareholder value (e.g., Pfau & Kay, 2002); however, an alternative view (Boudreau & Ramstad, 2003; McLinden, 1995) is that organizational decision makers are responsive to logical evidence that HR practices are done well and are logically related to key business objectives.

Third, ensure that there are adequate resources for the evaluation. Presumably, some thought has already been given to the financial and administrative resources required to carry out the evaluation. However, consider the impact on those involved in the program. Will participants or coaches be required to complete multiple and/or lengthy rating forms? Will 360-degree appraisals be given to peers or subordinates of participants? Who else will provide data and how much time will be required to do so? Make sure your expectations are realistic by discussing them with representatives of all who will be involved in data collection.

Fourth, anticipate obstacles you may encounter regarding either conducting the evaluation or presenting the findings. Will you have limited time or access to collect data from coaches or participants? Who will look bad if the results are not positive? Will organizational priorities toward evaluation change? Will the results of the study be met with skepticism? Listing these obstacles during planning may lead to better choices of whom to involve in planning and what level of rigor you should strive for in your research design.

Finally, take steps to anticipate when cultural, diversity, and language issues are relevant. This is another reason to involve a broad set of stakeholders at the outset of the evaluation process, as they may help you identify, for example, how those with different cultural values may perceive the process. Multinational organizations that provide coaching services in various languages need to consider at an early stage how they will gather data from those who speak the nondominant language.

Often, such a process requires translating surveys into a second language and then translating responses back into the original language for analysis and aggregation with other inputs. Disseminating the findings of the program evaluation at the conclusion may involve yet another translation process.

Step 2: Design the Process

After firmly grounding the evaluation, use your analysis and insights to create measures and design the data-gathering process.

First, determine who will provide the information you need. If you are interested in marketing, you will probably want comments from participants and bosses about the value of the program. If you wish to improve the program, coaches, participants, and their bosses can talk about logistics, program administration, communication, organizational support, and the process of coaching itself. If you want evidence of effectiveness, input from participants and bosses is often most compelling; you might even seek objective performance measures that are related to the coaching objectives. Depending on the specific coaching objectives, you may seek input from participants and bosses, as well as their peers, direct reports, and customers, depending on who is most likely to see the relevant behaviors. Organizational sponsors and other stakeholders should also be considered as sources of input.

Second, generate the specific questions and items they will respond to. It often pays to do some homework inside the organization to see whether you can borrow or adapt existing items that have proven their value. In some instances, the evaluation of coaching programs may be supplemented by using existing organizational data such as information on retention, measures of work unit performance, employee surveys, 360-degree evaluation, performance appraisals, or assessment centers.

If your purpose is to gather feedback to improve the coaching process, you might ask similar questions of all parties (recognizing that participants, coaches, bosses, and HR personnel have different needs and expectations and will therefore often comment on different aspects of the program): What recommendations do you have to improve the coaching process? What aspects of the process are most useful? What could be done to meet your needs more effectively? What environmental or organizational barriers, if any, were most detrimental to the process? To ensure a more comprehensive response, include a checklist of issues for each person to either rate or to consider as part of his or her open-ended responses (e.g., clarity of expectations, understanding of the process at the outset, quality and frequency of communication, quality of feedback, confidentiality, scheduling and logistics, length and frequency of meetings, access to other resources; see also Executive Coaching Forum, 2001; Peterson, 2002). For additional depth on the participant perspective, you might ask more specific open-ended questions (e.g., What is most or least useful about the coaching you are receiving? How did your coaching help you achieve better results at work?). Another alternative is to systematically guide them through aspects of the process by having them rate items such as the following:

- Administrative processes and support (e.g., scheduling, communications) have been smooth and responsive.
- My coach tailored the process to address my specific needs and situation.
- My coach has been positive, professional, and supportive.
- My coach understands our organization and its culture.
- My coach was effective in helping me *learn* new behaviors.
- My coach was effective in helping me *apply* new behaviors on the job.
- My manager provided clear, useful feedback.
- My manager was supportive of the coaching process.

To aid decision making or marketing, you may want to evaluate the extent to which participants have learned new skills or improved their performance. The simplest method is to ask about overall outcomes that apply to all participants, such as effectiveness on the job, potential for advancement, and general leadership abilities (e.g., To what extent did coaching improve this person's leadership capabilities?). You may also measure a common set of behaviors related to your organization's values or key learning behaviors, such as these: Actively seeks feedback from others; Takes appropriate risks in order to put learning into action.

Because coaching is individually customized, the most precise evaluation, as well as the most useful feedback for participants and coaches, will be gained by using customized items related to each person's own development priorities. In general, we recommend using both common items (for easy comparison) and unique items (for evaluating individual progress).[2] Representative items include the following:

- Focus your organization on efforts that add significant value to customers.
- Provide strong, decisive leadership despite incomplete and ambiguous information.
- Consider and address the impact of your plans and changes on other business units.

Finally, if marketing is the purpose, items can be included purely to gather information that might appeal to others:

- Overall, this is a high-quality program.
- I would recommend my coach/this process to others.
- The coaching I received was well worth the cost.

Your third decision is to choose the appropriate rating scales to be applied. For example, if the sole purpose of the preceding three items is marketing, it might be best to ask them in a yes/no format so you can say, "99% of participants would recommend this coaching program to others." If your purpose is decision making or improving the process, you need the finer grained information provided by a 5- or 7-point scale.[3]

For rating participants on the amount of improvement they show against their coaching objectives, we recommend choosing from the following rating scales:

■ *Current Effectiveness* (ranging from 1 = Not at all effective, 4 = Acceptable, 7 = Highly effective). This scale is most appropriate when you are able to compare pre-coaching and post-coaching ratings, or make comparisons involving a control group.

■ *Current Versus Desired Performance* (1 = Much less than desired performance, 4 = Equal to desired performance, 7 = Much better than desired performance). This scale may be combined with a current effectiveness scale and is most appropriate when giving feedback is a purpose for the evaluation.

■ *Level of Improvement* (1 = No improvement, 3 = Moderate improvement, 5 = Considerable improvement). This scale is useful for measuring the effectiveness of coaching and is extremely valuable when you are unable to collect baseline performance information (Peterson, 1993a). Because it does not clearly indicate whether a person's performance meets expectations, even when substantial change has occurred, it is most useful in conjunction with one of the other two scales.

Fourth, decide how to collect your data: online, paper-and-pencil surveys, interviews, or through some combination. This, as with all other decisions, should be driven primarily by your purpose. If you are conducting frequent surveys from multiple rater perspectives in widely dispersed areas, online is often easiest. If you are seeking feedback to improve the program design, open-ended questions are most useful. They can be delivered through any vehicle, but interviews allow for deeper discussion and additional follow-up questions. Whatever method you use, be sure to include an explanation of how the information will be used as well as the level of confidentiality and anonymity that you will provide to those responding to your questions.

Fifth, consider the design that will give you the most compelling information for your purpose. For example, pure marketing efforts have less need of rigorous design (pre-/posttesting, comparison groups, item-level controls, etc.) than do evaluation efforts driven by decision making. The point here is to start by carefully designing a thorough and flexible approach to give you all the information you are likely to need. As resource or business factors limit your options, compromises can be made. However, by beginning with an optimal design, you are more assured that the final design will address potential criticisms of the evaluation results.

Recommendations

With such an array of options in front of you, the ideal design is not always clear. Based on our experience balancing typical evaluation needs with the practical realities of data gathering in most organizations, we can offer the following recommendations.

■ For feedback to coaches and participants, and to facilitate open communication with the participant's boss, HR, and other organizational sponsors, use a pre- and post-coaching rating of 8–10 specific learning objectives for each participant. At Time 1, use either the Current Effectiveness or the Current Versus Desired Performance scale. At Time 2 (or at any point during the coaching process), use the

	Strongly Disagree		Neutral		Strongly Agree
1. Overall, this is a high-quality program.	1	2	3	4	5
2. Administrative processes and support (e.g., scheduling and communications) have been smooth and responsive.	1	2	3	4	5
3. My coach and coaching program addressed my specific needs and situation.	1	2	3	4	5
4. My coach has been positive, professional, and supportive.	1	2	3	4	5
5. My coach was effective in helping me *learn* new behaviors.	1	2	3	4	5
6. My coach was effective in helping me *apply* new behaviors on the job.	1	2	3	4	5
7. I would recommend this coaching process to others.	1	2	3	4	5

As a result of my coaching . . .

8. I have a better understanding of my impact on others.	1	2	3	4	5
9. I am more effective on the job.	1	2	3	4	5
10. Others have commented on my increased effectiveness on the job.	1	2	3	4	5
11. I have improved in critical skills that increase my potential for advancement.	1	2	3	4	5
12. I am a more effective leader and/or influencer.	1	2	3	4	5
13. I learned how to keep learning and improving in the future.	1	2	3	4	5

Open-ended questions

14. What are the two or three most important things you personally learned?

15. What were the most helpful and beneficial parts of the program for you?

16. What two things would have made this a more positive and valuable experience for you?

17. Other comments or suggestions:

Figure 13.2 Participant Survey

same scale again plus the Level of Improvement scale. The Time 2 measure can be used even when you do not have Time 1 data.

■ For a simple, all-purpose instrument, use a format similar to Figure 13.2 for participants and Figure 13.3 for the boss and other organizational sponsors. The instrument is designed to be administered at the completion of coaching. Note the range of items and how they can be used to address multiple decision making, improvement, and marketing purposes.

■ If you are interested in ROI and the business impact of the coaching, it is essential to gather information from the boss and/or other organizational sponsors at the outset of coaching (as discussed in the ROI section above) regarding their desired outcomes for the coaching and what they perceive the *value* of those outcomes to be.

	Strongly Disagree		Neutral		Strongly Agree
1. Overall, this is a high-quality program.	1	2	3	4	5
2. The coach and coaching program have addressed our specific needs and situation.	1	2	3	4	5
3. The coach has kept me informed and involved.	1	2	3	4	5
4. The coach was effective in helping this person *learn* new behaviors.	1	2	3	4	5
5. The coach was effective in helping this person *apply* new behaviors on the job.	1	2	3	4	5
6. The coach gave me tactics on how to support this person's development.	1	2	3	4	5
7. I would recommend this coaching service to others.	1	2	3	4	5
8. Considering all aspects of performance, this person is very *effective* on the job.	1	2	3	4	5
9. The coaching and training *contributed* significantly to this person's improved effectiveness on the job.	1	2	3	4	5
10. Overall, I am very satisfied with this person's *progress* in the coaching program.	1	2	3	4	5
11. I am very satisfied with this person's *effort* in the coaching program.	1	2	3	4	5
12. At this point, should an opening appear, this person will be likely to *advance* within our organization.	1	2	3	4	5
13. Others in the organization now feel more confident giving this person's team challenging or difficult assignments.	1	2	3	4	5
14. This person's team is more effective at leading and influencing within the organization.	1	2	3	4	5
15. This person's team gets more done.	1	2	3	4	5
16. Morale in this person's team has improved.	1	2	3	4	5

Open-ended questions

17. What are the two or three changes this person has made that have most impacted your organization?

18. *How* have these changes impacted your organization?

19. What two things would have made this a more positive and valuable experience for you?

20. Other comments or suggestions:

Figure 13.3. Organizational Sponsor Survey

Step 3: Implement the Process

At this point, you implement the decisions made at Step 2—for example, collecting the data using online, e-mail, paper-and-pencil, or interview methods.

It is important to remember here that lack of time by raters is one of the primary obstacles in conducting evaluations (Kraiger, 2002). If you involved stakeholders in the planning process at Step 1, it may be easier to elicit responses from your raters. Recognize that getting timely and accurate ratings will be one of the greatest challenges you face in the evaluation process.

There are four suggestions for improving both response rates and the quality of the data. The first, already stated, is to gain buy-in by involving as many potential raters as possible in the planning of the evaluation and the design of the rating form. For example, you can circulate a draft of the instrument to participants, coaches, and bosses and ask for suggestions for improving the form or adding items. Second, if possible, dedicate time for completing ratings, such as reserving the last 15 minutes of a staff meeting for this purpose. Third, if raters complete the forms on their own time, plan to send one or two follow-ups reminding raters to return the forms by the given deadline. It is usually best to give short initial deadlines, no more than 5 to 7 business days. The longer the initial deadline, the more likely raters are to delay in completing their ratings.

Fourth, it is important to maintain rater confidentiality and protect the integrity of the data throughout the rating process. This is actually a long-term process to build trust and should be managed before, during, and after the data collection process. It is important to answer questions honestly about the purpose for the evaluation and the use of the ratings. It is also critical to safeguard the data and ensure that the data are used for nothing other than the intended purpose.

Step 4: Analyze the Data

This step involves analyzing your data to address the questions raised by the purpose of your study. A tutorial on the techniques of statistical analysis is beyond the scope of this chapter. Instead, our comments are directed at the goals of analysis.

One of the major issues confounding interpretation of results is sample attrition, as it is unlikely that all participants who started the coaching program will have completed it. Calculating the percentage of participants who dropped out can be informative. A dropout rate that is particularly high or low may say a lot about the value participants see in the program or help you identify other obstacles that need to be removed. From the perspective of statistical validity (i.e., can we believe the results?), it is useful to compare participants who dropped out and participants who stayed. Do program dropouts differ from those who stayed in terms of age, gender, ethnicity, program objectives, or type of coach? Occasionally participants will choose to leave the organization as a result of their coaching, which in cases of poor fit may be viewed as a successful outcome of the coaching process. The results of these comparisons may not only address the generalizability of the primary results but provide useful feedback on who is more or less likely to benefit from coaching in your organization, desirable characteristics of coaches, how participants and coaches are paired, or how the coaching can be improved.

For obvious reasons, the analyses you conduct will be based on the measures and design you selected at Step 2. More generally, these analyses should be planned with

consideration to the issues of impact, effectiveness, and efficiency addressed in the previous section. To evaluate impact, determine whether the coaching program itself was consistent with the strategic objectives of the organization. Did specific coaching objectives support these strategic objectives? These questions can be answered by rational analysis or judgments by stakeholders. To evaluate effectiveness, conduct statistical analyses of your measures to determine whether program objectives were met and whether coaches improved over the duration of the program. Finally, to evaluate efficiency, consider whether the observed benefits were worth the investment by the organization in coaching. Consider too whether program efficiency can be improved by streamlining some aspects.

A detailed discussion of how to collect data to conduct cost-benefit analyses is beyond the scope of this chapter, but readers interested in detailed presentations are referred to Cascio (1982), Phillips (1997), or Tracey (1992). McGovern et al. (2001) also present a case specific to coaching interventions. A comparison of the costs and benefits of coaching is the most straightforward when both are expressed in the same units of measurement.[4]

According to Tracey (1992), cost and benefit data may be collected from three sources: corporate records and reports, HR department records, and directed accounting surveys and studies. In typical studies, data are collected on both direct and indirect costs. Direct costs include travel expenses, out-of-pocket expenses by participants, and personnel costs associated with the development and maintenance of the coaching program. Indirect costs are items of expense that are not associated with the specific coaching program but are common to multiple HR activities. Examples include fringe benefits (of administrative staff) and overhead.

Note that in contrast to formal training programs, both the direct and indirect costs of coaching programs will be relatively low. However, it is also important to consider opportunity costs, the costs of foregoing other services by participation in the coaching program. Specifically, each hour spent by the participant and coach on coaching is an hour not spent on contributing directly to the primary operation of the business. Thus, it may be appropriate to calculate or estimate time spent in coaching and the hourly salary of participants as costs associated with the program.

Following traditional ROI models, these accumulated costs can be compared to some estimate of the total benefit of the program. Using the same sources as above, it may be possible to estimate reduction in operational costs or errors in job performance, greater individual or unit productivity, or customer retention. For example, after coaching on time management, a manager estimates that she can accomplish her work 15% more efficiently than in the past. Because the investment of her time in coaching was counted as a cost, it is appropriate to count the increased efficiency (.15 times her hourly salary times the length of the intervention) as a benefit.

There are several strategies for comparing costs and benefits including calculating the ratio of benefits to savings (see Cascio, 1982; Phillips, 1997; and Tracey, 1992, for more details). Note too that by using the impact/effectiveness/efficiency approach of Boudreau and Ramstad (2003), it is less important to express the cost effectiveness of the coaching program as a single number than it is to show that coaching creates tangible outcomes, linked to important business goals, while making efficient use of corporate resources.

Step 5: Present the Findings

Finally, present your findings to the relevant audience. Note that you might even choose to disseminate the results of the evaluation more broadly, so that all raters can understand the organizational value of their responses. Regardless of how you present the findings, whether in person or in a written document, be sure the context is clear. On the one hand, this involves linking the results back to the original purposes of the evaluation and to the specific questions that were targeted at the outset of the process. On the other hand, the context requires that the results be interpreted in light of who was enrolled in coaching for what purpose. A coaching program offered to high potentials to use as they wish will undoubtedly generate different findings from a coaching program aimed at potential derailers who are given few choices about their participation.

Other points to sharing your findings are provided in Figure 13.1, so we do not go into additional detail here other than to stress the importance of additional dialog around the results. Present your findings and then talk with your audiences about their reactions. Use your new insights to identify top priorities for action. You may even go back with a clearer or revised sense of the purpose of evaluation. The evaluation process is an ongoing task that benefits from periodic review to ensure that it is serving current purposes well.

Final Comments

Evaluation is primarily about making a *credible case* to support important business priorities. Managers are looking for practical information they can use to achieve organizational objectives and solve problems. The process you design for evaluating coaching must be based on an understanding of the business needs, the types of decisions that will be made, and what your various audiences will find most useful. By considering the need for evaluation early in any coaching process, you can facilitate learning and continuous improvement, and enhance your own value and credibility with your organization's leadership team.

Notes

1. Compare these results to Bloom's (1984) "two-sigma" difference between one-on-one tutoring and classroom learning techniques for children.

2. Although customized learning objectives lead to the best data for the individual, they present a potential secondary problem in the next stage of evaluation—data analysis. There are at least three choices in handling these disparate (i.e., individually customized) pieces of information. (1) Include both individual items and common items in the rating process, using the individual items exclusively for individual feedback and the common items for program-level evaluation. (2) Evaluate each person's progress separately, based on his or her individual ratings, and then combine that information to gauge the overall program. For example, you might conclude that 27 participants showed significant improvement, 12

moderate improvement, 5 no change, and 2 an actual decline. (3) Average all items together as if they were equivalent. Peterson (1993b) concluded, after multiple examinations of the data, that option three is appropriate if all of the coaching objectives are written to express a uniform level of competence. For example, it is *not* reasonable to assume that a rating on "Demonstrates world-class strategic thinking abilities without exception" can be averaged with "Demonstrate strategic thinking in key decisions." An average rating on the former might actually be equivalent to a top rating on the latter.

3. As you will note in the following scales, we recommend using 7-point scales with a range from highly positive to highly negative ratings, such as in the first two scales listed here. When one end point is zero, such as the "No improvement" anchor on the Level of Improvement scale, a 5-point option is sufficient.

4. Alternatively, costs and benefits may be expressed in different metrics, leading to statements such as "a $20,000 investment in executive coaching is related to a 4% decrease in employee turnover or 5% increase in customer satisfaction."

References

Birkeland, S., Davis, B., Goff, M., Campbell, J., & Duke, C. (1997). *AMOCO individual coaching study.* Unpublished research report, Personnel Decisions International. Minneapolis, MN.

Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 4–16.

Boudreau, J. W., & Ramstad, P. M. (2003). Strategic industrial and organizational psychology and the role of utility analysis models. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski,(Eds.), *Handbook of psychology, Vol. 12: Industrial and organizational psychology.* Hoboken, NJ: Wiley.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71,* 232–246.

Cascio, W. F. (1982). *Costing human resources: The financial impact of behavior in organizations.* Belmont, CA: Wadsworth.

Davis, B. L., & Petchenik, L. (1998). *Measuring the value of coaching at AMOCO.* Presented at The Coaching and Mentoring Conference, Washington, DC: Linkage, Inc.

Gegner, C. (1997). *Coaching: Theory and practice.* Unpublished master's thesis, University of San Francisco, San Francisco, CA.

Haccoun, R. R., & Hamtiaux, T. (1994). Optimizing knowledge tests for inferring acquisition levels in single group training evaluation designs: The internal referencing strategy. *Personnel Psychology, 47,* 593–604.

Kampa, S., & White, R. P. (2002). The effectiveness of executive coaching: What we know and what we still need to know. In R. L. Lowman (Ed.), *Handbook of organizational consulting psychology* (pp. 139–158). San Francisco: Jossey-Bass.

Kampa-Kokesch, S. (2001). *Executive coaching as an individually tailored consultation intervention: Does it increase leadership?* Unpublished doctoral dissertation, Western Michigan University, Kalamazoo, MI.

Kampa-Kokesch, S., & Anderson, M. Z. (2001). Executive coaching: A comprehensive review of the literature. *Consulting Psychology Journal, 53*(4), 205–228.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and maintaining effective training and development: State-of-the-art lessons for practice* (pp. 331–375). San Francisco: Jossey-Bass.

Martineau, J. W., & Preskill, H. (2002). Evaluating the impact of organization development interventions. In J. Waclawski & A. H. Church (Eds.), *Organization development: A data driven approach to organization change* (pp. 286–301). San Francisco: Jossey-Bass.

McGovern, J., Lindemann, M., Vergara, M., Murphy, S., Barker, L., & Warrenfeltz, R. (2001). Maximizing the impact of executive coaching: Behavioral change, organizational outcomes, and return on investment. *The Manchester Review, 6*(1), 1–9.

McLinden, D. J. (1995). Proof, evidence, and complexity: Understanding the impact of training and development in business. *Performance Improvement Quarterly, 8*(3), 3–18.

Peterson, D. B. (1993a). *Measuring change: A psychometric approach to evaluating individual coaching outcomes.* Presented at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco.

Peterson, D. B. (1993b). *Skill learning and behavior change in an individually tailored management coaching program.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Peterson, D. B. (2002). Management development: Coaching and mentoring programs. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 160–191). San Francisco: Jossey-Bass.

Peterson, D. B., & Hicks, M. D. (1999, February). *The art and practice of executive coaching.* Presented at the annual conference of the Society of Consulting Psychology, Phoenix, AZ.

Pfau, B., & Kay, I. (2002). Playing the training game and losing. *HR Magazine, 47*(7), 49–54.

Phillips, J. J. (1997). *Return on investment in training and performance improvement programs.* Houston: Gulf Publishing.

Thompson, A. D., Jr. (1986). *A formative evaluation of an individualized coaching program for business managers and professionals.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Tracey, W. R. (1992). *Designing training and development systems* (3rd ed.). New York: Amacom.

West, L., & Milan, M. (2001). *The reflecting glass: Professional coaching for leadership development.* Basingstoke, Hampshire, Great Britain: Palgrave.

Suggested Readings in Coaching

Executive Coaching Forum. (2001). *The executive coaching handbook: Principles and guidelines for a successful coaching partnership* (2nd ed.). Author. (Available at *www.theexecutive-coachingforum.com*)

Fitzgerald, C., & Berger, J. G. (Eds.). (2002). *Executive coaching: Practices and perspectives.* Palo Alto, CA: Davies-Black.

Hargrove, R. A. (1995). *Masterful coaching.* San Diego: Pfeiffer.

Peterson, D. B. (2002). Management development: Coaching and mentoring programs. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 160–191). San Francisco: Jossey-Bass.

Peterson, D. B., & Hicks, M. D. (1996). *Leader as coach: Strategies for coaching and developing others.* Minneapolis, MN: Personnel Decisions International.