# Chapter 6: Evaluating coaching programs

Elza Mylona, PhD, MBA; Richard Van Eck, PhD

### Take home points

1. Program evaluation depends on the goals; the clearer they are, the more effective the program will be.
2. Evaluation models are powerful tools, but expertise in using them is critical.
3. Evaluation begins before, and continues after, the intervention itself.
4. Success criteria should be defined and stakeholders' participation ensured before choosing evaluation measures.

## Impact of coaching

Coaching has the potential to facilitate developmental changes for individuals and organizations. In business settings, coaching programs offer an accessible and cost-effective option because they can help executives adapt to new roles or environments on the job.[1] Coaching is also considered one of the most powerful methods for developing soft skills.[1] It may also be used as a remedy for poor performance or to help navigate extreme organizational changes.[2]

While there is evidence that coaching can be effective, there remains a need for more rigorous evaluation.[1-3] This chapter will outline some of the key aspects of effective evaluation of coaching programs.

## Program evaluation

Program evaluation is the systematic collection and analysis of information related to the design, implementation and outcomes of a program for the purpose of monitoring and improving its quality and effectiveness.[4] Information is collected systematically and deliberately, following the same rigorous methods applied in other types of research

### Vignette

*Vick Reardon has been hired as the dean for faculty development at a medical school. His dean has asked him to evaluate a coaching program where faculty who are great teachers are chosen to serve as coaches.*

*The dean wants to know if the program has produced meaningful outcomes to justify the stipends she pays to fund it. She wants documented, "evidence-based results."*

*Vick finds little documentation about how the program was designed or what problem(s) it was intended to solve, let alone any data regarding its efficacy. An analysis of instructor rating forms completed by students each semester shows no statistically significant differences before or after the implementation of the program.*

### Thought questions:

If you were in Vick's place, what would you tell the dean about:

1. The timing of the evaluation planning in relation to the program,
2. The quality of the program as designed and implemented, and
3. Whether it is possible to answer her evaluation question and, if so, what would be required?

to identify whether the program is effective and what contributes to the success of the program.[5] Evaluations can be done for many purposes, including to demonstrate program effectiveness to funders, to improve the implementation and effectiveness of programs, to manage limited resources, to justify program funding, to satisfy ethical responsibility to clients, or to document program development and activities to help ensure successful replication. Regardless of the purpose, all program evaluations share six activities in common.

## Table 1. Typical program evaluation activities

| |
|---|
| Posing questions about the program |
| Setting standards for effectiveness |
| Designing the evaluation and selecting the participants |
| Collecting data |
| Analyzing data |
| Reporting results |

## Posing evaluation questions

Program evaluation starts with one or more questions, sometimes simple and easy to answer but more often complex. A good place to start in developing evaluation questions is by asking questions such as, Why do we want to evaluate this program? How will we use the results? What are the evaluation goals of the stakeholders? What will tell us whether the program is performing as designed?

## Table 2. Common program evaluation questions

| |
|---|
| To what extent did the program achieve its goals and objectives? |
| For which individuals or groups was the program more effective? |
| What are the characteristics of the individuals or groups who participated? |
| How enduring were the effects? |
| Which features (activities, settings, management strategies) of the program were most effective? |
| To what extent are the program's objectives and activities applicable to other settings/institutions? |
| What are the relationships between the cost and the effects of the program? |
| To what extent did changes in social, financial, political, etc., circumstances influence the program's support and outcomes? |

## Setting standards for effectiveness

Program evaluations are concerned with evidence of program effectiveness. Programs are often measured against particular standards; the more specific the standards are, the easier they are to measure. Standards can be established by reviewing other comparable programs, reviewing the literature, or relying on the consensus of experts. The challenge is to identify standards that are credible as well as appropriate and possible to measure.

## Designing the evaluation and selecting the participants

Ideally, program evaluation should be considered concurrently with design of the coaching intervention, as the one can influence the other. Standard evaluation designs include comparing one group's performance over time (when all participants have received the same training) or comparing two groups at one or more times (when some participants receive a different program or no program at all). Some of the general questions to consider at this point may include the following: How many measurements should be made? When should the measurements be made? How should the groups or individuals be chosen?

## Collecting data

The process of collecting and measuring information on variables of interest enables one to answer stated research questions, test hypotheses, and evaluate outcomes. Some general principles to consider include finding out what data are already being collected, keeping the evaluation questions front and center to ensure only the necessary data are collected, collecting data from more than one source for each question, and collecting a mix of quantitative and qualitative data. Quantitative data are useful for discovering the magnitude of a phenomenon (e.g., outcomes, barriers, facilitators). Qualitative are useful to better understand the phenomenon (e.g., who benefits most from a program, what additional support is needed to improve outcomes).

## Analyzing data and reporting results

The method of analysis depends on the evaluation questions and the standards selected. It is important to consider both the practical significance (how much impact the change represents) as well as statistical significance (whether the change is detectable by statistical measures). A lack of statistical significance may indicate that outcome measures were too ambitious or the desired behavioral change may take longer to emerge. Conversely, some "nonsignificant" findings may end up being useful for understanding or modifying the program. Interpreting results and drawing conclusions from program evaluations can be challenging. The involvement of stakeholders in reviewing findings and preliminary conclusions prior to writing a formal report is highly recommended.

## Common types of program evaluations

Most program evaluations focus on outcomes, goals, or processes, as discussed below.

### Outcome-based evaluations

Outcome evaluation is the most commonly requested evaluation by accrediting bodies. It assesses whether the program is producing the desired change. It focuses on what changed for program participants and how much difference those changes in turn made for them and the institution. These types of questions are among the most difficult to answer because it is not always possible to isolate the results of a program from other factors. Careful specification and alignment of program goals, outcome measures, and evaluation instruments and procedures is required. Siegfried[6] argues that we should distinguish between 1) the ***general measures*** of success that include areas such as goal attainment, satisfaction of the participants involved, affect change, well-being, and life satisfaction and 2) the ***specific measures (outcomes)***, which will depend on the coaching intervention

proposed or the problem being addressed, such as improvement of clinical skills, more effective coping with stress, or improvement of academic performance and attainment of competencies. Outcome-based questions may include questions like these: Did the coaching program succeed in helping students transition to residency? Was the program more successful with certain groups of students or specialties than with others? What aspects of the program did participants find gave the greatest benefit?

### Goal-based evaluations

These types of evaluation look at the extent to which the coaching program has met its predetermined goals or objectives. They do not evaluate whether the goals themselves are valid, nor whether the measures of effectiveness being used are effective. Goals-based evaluation questions include questions such as these: How were the goals of the program established? Was the process effective? If not, why? What is the status of the program's progress toward achieving its goals? Will the goals be achieved based on the time line established? If not, why? Do the people involved in the program have adequate recourses (time, training, facilities, and budget) to achieve the goals? How should priorities be changed to ensure completion of goals?

### Process-based evaluations

Process-based evaluations focus on the program's activities rather than its outcomes. Activities may include the types and quantities of services delivered, the beneficiaries of those services, the resources used to deliver the services, the practical problems encountered, and the ways such problems were resolved. Process evaluations are similar to the concept of implementation fidelity studies, in which one measures how well the intervention was implemented and allow others to replicate the programs in other settings and contexts. In this sense, many program evaluations will include some form of process-based evaluation. Process evaluation questions may include these: Was the coaching program successful? If so, how and why? What were the kinds of problems encountered in delivering the program? Was the program well managed? Were participants trained or educated to the right level for the program? Was there adequate support for the program?

## Common evaluation models

One of the best ways to come to grips with the evaluation process is to become an expert with one or more models for evaluation. These models serve as heuristics to scaffold the evaluation in ways that make the process more manageable.

## Kirkpatrick's model

Kirkpatrick's model has been widely used for conducting outcome evaluations of training programs.[7] This model supports the gathering of data to assess four "levels" of program outcomes.

## Table 3. Kirkpatrick's evaluation model

| Levels | Key features | Cautions and caveats |
|---|---|---|
| Level 1: Reaction/ satisfaction | Evaluates how participants perceive the training. | Measures perception, not learning. |
| Level 2: Learning | Evaluates actual learning. | Should be measured before and after training. |
| Level 3: Behavior | Whether what was learned (L2) actually transfers to workplace.<br>Focus is on behavioral changes<br>Observations/interviews with employees/supervisors are common. | May take weeks or months to manifest/assess.<br>Environmental pressures may prevent change. |
| Level 4: Results | Impact of changed behavior on the organization. | Requires significant resources/ long-term commitment.<br>Best reserved for mission-critical outcomes and long-term programs. |

## The logic model

This model represents a narrative or graphic depiction of real-life processes that communicate the underlying assumptions upon which a specific activity is expected to lead to a specific result.[8] It describes logical linkages among program resources, activities, outputs, and audience and short, intermediate, and long-term outcomes. It can be very linear in its approach to planning and evaluation, which may oversimplify the complexity of training interventions.

**Table 4. The logic model**

| Component | Key features | Cautions and caveats |
|---|---|---|
| Inputs | Everything invested in the program (e.g., knowledge, skills, expertise, funding resources, time, facilities, equipment, technology). | Does not measure actual learning.<br>Creating inventory of "inputs" helps stakeholders understand scope and commitment. |
| Activities | Explicit ordering of activities expected to take place.<br>Specific sequence of activities in the model.<br>Results of measurement used to make adjustments to activities or outputs. | Sequence should illustrate interdependency of activities.<br>Each activity requires its own measurement criteria. |
| Outputs | The things done as a result of the activities (e.g., develop curriculum or other resources, conduct training, recruit coaches, deliver workshops) and the activities of people impacted. | Every activity must have at least one output.<br>One output may have more than one activity associated.<br>One activity may have multiple outputs. |
| Outcomes | Outcomes answer this: "What happened as a result of the program?"<br>Short term (e.g., demonstration of knowledge, skills, attitudes, awareness, opinions, motivations).<br>Intermediate term (e.g., changes in participant practice or behaviors exhibited by participants or institution, polices adopted by the institution).<br>Long term (e.g., changes in faculty compensation plan to support coaching, improved teamwork and collaboration). | Short term = Kirkpatrick's Level 2.<br>Intermediate term = Kirkpatrick's Level 3.<br>Long term = Kirkpatrick's Level 4. |

## In review

The scenario presented in the beginning, while perhaps extreme, is indicative of the major threats faced when designing and evaluating coaching programs.

Vick can tell the dean that evaluation planning should have been done at the same time as the planning of the program. Evaluation shapes implementation (e.g., documentation and data collection procedures, benchmarks for performance), and program needs determine evaluation methods (e.g., models, measurement). Without knowing the goals of the program and the measurement criteria envisioned, Vick and the dean cannot make definitive statements about program quality. Observation would suggest, however, that participants are not aware of criteria, that implementation fidelity is therefore weak, and that efficacy data have not been designed nor systematically collected. Furthermore, it is not currently possible to answer the dean's evaluation question. Evaluation components will have to be recreated, measurement tools designed, implementation guidelines generated and applied, questions formulated, and an evaluation model selected and implemented over at least the next year.

## Conclusion

While program evaluations may seem complicated, expensive, or even overwhelming, they are critical for improving programs. Evaluations can provide process data on the successes and challenges of early implementation or, for more mature programs, can provide outcome data on program participants. The information obtained can help to target program resources in the most cost-efficient way. The key is to understand the questions to be answered and adopt an appropriate, familiar model while conducting evaluation concurrently with program design. Then how well the program is being implemented (process) can be measured as well as its impact on the organization (outcomes or goals).

## References


1. Leonard-Cross E. Developmental coaching: Business benefit—fact or fad? An evaluative study to explore the impact of coaching on the workplace. *International Coaching Psychology Review*. 2010;5 (1), 36–48.
2. McGovern J, Lindemann M, Vergara MA, Murphy S, Barker L, Warrenfeltz R. Maximizing the impact of executive coaching: Behavioral change, organizational outcomes and return on investment. *A Journal for People and Organizations in Transition. The Manchester Review*. 2001;6 (1), 1–9.
3. Fillery-Travis A, Lane, D. Does coaching work or are we asking the wrong question? *International Coaching Psychology Review*. 2006;1 (1), 23–36.
4. ACGME. Accreditation council for graduate medical education: glossary of terms. *Accreditation Council for Graduate Medical Educat*ion. 2013. Available at: https://www.acgme.org/Portals/0/PDFs/ab_ACGMEglossary.pdf
5. Frye AW, Hemmer PA. Program evaluation models and related theories. AMEE Guide No. 68. 2012. Available at: http://dx.doi.org/10.3109/0142159X.2012.668637
6. Siegfried G. Advances in research on coaching outcomes. *International Coaching Psychology Review.* 2008;2 (3), 222–249.
7. Kirkpatrick D. Revisiting Kirkpatrick's four-level model. *Training and Development.* 1996;50 (1), 54–58.
8. McCawley PF. *The logic model for program planning and evaluation*. University of Idaho Extension. 2002. Available at: https://www.cals.uidaho.edu/edcomm/pdf/CIS/CIS1097.pdf