• Introduction •

Performance Improvement and Evaluation

You may have been to a performance theater to attend a ballet or modern dance performance, or listen to an orchestra playing classical music. Regardless of the types of performing arts that you have attended, the artists probably went through years of training and practice before they performed in front of the audience. You do not see the behaviors that the artists exhibited during their training process; you only see their performance outcome and applaud for their accomplishments at the end of their performance.

Similarly, performance in the workplace consists of workers' behaviors and their accomplishments (Chyung, 2008; Gilbert, 2007). Workers' behaviors enable their accomplishments. However, organizations value not just workers' behaviors but, more importantly, their accomplishments. For that reason, the phrase *human performance* (or simply *performance*) in the workplace often refers to workers' accomplishments rather than their work behaviors. Competent workers likely produce accomplishments that are valued by not only themselves but also their organization and society. However, workers can become competent and successful only if they are in an appropriate environment (Rummler & Brache, 2012). Thus, performance improvement in the workplace involves finding ways to develop workers' capacity and change environmental factors that contribute to producing valuable performance outcomes.

Human performance improvement (HPI) is a transdisciplinary field of practice that promotes the use of systematic and systemic approaches to engineering work processes (behaviors) and producing desirable outcomes (accomplishments). HPI practitioners find areas to be changed, implement cost-effective solutions, and produce results valuable to the workers, their organization, and the society. Here, we should pay attention to a couple of phrases: a transdisciplinary field of professional practice and systematic and systemic approaches.

HPI is a transdisciplinary field of practice in that it synthesizes knowledge adopted from various disciplines such as psychology, instructional technology, human resource development, management, organizational development, and evaluation. Its synthesized knowledge is applied to various disciplines. Compared to specific disciplines such as nursing or hospitality management (where professionals are educated to acquire a specific branch of knowledge and skills and hold specific job titles with established job responsibilities), HPI is currently a field of professional practice that is not limited to a certain discipline. The existence and development of HPI as a field of practice depends on its assimilation into other disciplines. HPI practitioners work within various industries and organizations in numerous capacities such as instructional designers, trainers, workforce development specialists, human resource specialists, performance improvement professionals, or consultants. HPI can also be an

additional professional role or a responsibility that professionals add to their existing job position. For example, nurses or hotel managers may also be educated to apply HPI principles to their daily practice.

Regardless of the type of industries and organizations where HPI practitioners work, they use systematic and systemic approaches to workforce development and performance improvement. **Systematic approaches** follow effective and efficient logical steps that lead to intended outcomes. **Systemic approaches** consider various factors both within and outside the immediate system that may influence the HPI process and outcomes positively or negatively. HPI practice requires a combination of both systematic and systemic approaches, as one often complements the other.

It should be clear by now that the term *HPI practitioner* is not a job title or position, but rather refers to any professional who applies HPI principles (such as systematic and systemic approaches to performance improvement) to any field or industry (i.e., transdisciplinary capacity). The systematic and systemic process of HPI practice is depicted in the human performance technology model (Van Tiem, Moseley, & Dessinger, 2012), comprised of several main phases:

1. Performance analysis, including organizational analysis, environmental analysis, gap analysis, and cause analysis

2. Intervention selection, design, and development

3. Intervention implementation and maintenance

4. Evaluation

5. Change management (which facilitates all previously listed phases)

Each of the five HPI phases is equally important. Evaluation is an essential phase, as it receives information from, and provides information to, the other phases. The information exchange that evaluation facilitates is often critical to the success of the other HPI phases. For example, existing evaluation data may be fed into a performance analysis of related issues, be considered when selecting appropriate solutions, or be used as a benchmark while monitoring performance improvement. Evaluation is conducted not just to improve performance outcomes but to continuously improve the performance improvement process itself.

What Is Evaluation?

So, what exactly is **evaluation**, and how does evaluation contribute to HPI? Although evaluation likely has been part of people's daily lives since early human history, evaluation as a profession emerged only a few decades ago (Stufflebeam & Shinkfield, 2007). As Michael Scriven puts it, "Evaluation is a new discipline but an ancient practice" (Scriven, 1991b, p. 3).

Ralph Tyler's work on the Eight-Year Study in the United States during the 1930s and 1940s is known as one of the pioneering works that facilitated the development of evaluation practice. During the study, new curriculum programs were evaluated to see if they produced expected outcomes as stated in the programs' objectives (Tyler, 1986). Evaluation—a.k.a. program evaluation—emerged as a field of professional practice in

the 1960s and 1970s to show accountability when a lot of federal monies were invested to develop and support government programs (Patton, 2008). Evaluation scholars and visionaries at that time also advocated using evaluation as a way to make continuous improvement on programs. Some of the fundamental principles and practices of evaluation such as summative evaluation (seeking accountability) and formative evaluation (seeking improvement) were developed at that time (Scriven, 1967).

The current *Merriam-Webster* dictionary definition of *evaluation* is "to judge the value or condition of (someone or something) in a careful and thoughtful way" ("Evaluation," n.d.). More specifically, Michael Scriven (2013) defines evaluation as "the process of determination of merit, worth, or significance" of someone or something, and is completed when the "value" is declared (p. 3). In fact, according to the *Concise Oxford Dictionary of Etymology,* the word *evaluate* originated from a French word, *évaluer; é-* means 'ex-' or 'out' (i.e., to find) and *valuer* means 'value' ("Evaluate," n.d.). Thus, to evaluate is to find the value of someone or something.

What Is Not Evaluation?

(Adapted from Chyung, 2015)[a]

Evaluation involves value assignment. It means evaluation is not the same as measurement because **measurement** does not include a value judgment. Consider measurement scales such as weight scales or measuring spoons. After you measured someone's weight to be 130 pounds, the weight itself does not tell you if the person is underweight, normal, or overweight. To know it, you need to calculate a body mass index (BMI) and compare it against a set of criteria that define underweight, normal, and overweight conditions. A weight of 130 pounds is normal for a 5'6" tall female adult, but it will put a 4' tall child in an obese category. Similarly, a tablespoon of sugar (measurement) does not tell you if it is too much or too little, until you put it in a context and evaluate it.

MEASUREMENT, ASSESSMENT, AND EVALUATION

Now, consider other measurement methods commonly used in the HPI context such as **survey (self-administered)**, **interview**, **focus group**, **observation**, **extant data review**, and **test**. These are data collection methods that measure people's thoughts, attitudes, or behaviors. Collecting data with these methods is not the same as conducting an evaluation. For example, administering a survey questionnaire to collect employees' attitudes toward their organization's culture is a measurement, not an evaluation. To *evaluate* whether employees have positive or negative attitudes toward their organization's culture, you need to analyze the survey data against standards or criteria (a.k.a. rubrics).

Another term, **assessment**, especially a test form of assessment, is sometimes used to mean evaluation. However, assessments are not evaluations either, unless the assessed results are given a value. A test score of 90 may represent an excellent performance if 60 is set as a passing score. However, it can be a failing score if a score of 95 is required. Think about personality assessments. These assessments are clearly not evaluations, since the assessment results only identify your dominant personality type but do not tell you whether you have a good or bad personality.

---

[a]The content of this section is from the following article with some modifications to make it fit within the chapter content: Chyung, S. Y. (2015). Foundational concepts for conducting program evaluations. *Performance Improvement Quarterly, 27*(4), 77–96. doi:10.1002/piq.21181

Thus, measurement and assessment are not the same as evaluation. However, you will likely perform measurements and assessments while conducting evaluations.

How Does Evaluation Compare With Research?

(Adapted from Chyung, 2015)[a]

To better understand evaluation, it is also important to understand similarities and differences between evaluation and research. First, **research**—especially social science research involving human subjects (which is the usual type of research conducted in the HPI context)—refers to a systematic data collection process concerning a given group of people in order to produce new knowledge about certain phenomena associated with the group. The new knowledge is either generalizable to its population (in quantitative research) or transferable to the understanding of similar groups or contexts (in qualitative research). Some research projects may employ both quantitative and qualitative research methods.

**SIMILARITIES AND DIFFERENCES BETWEEN EVALUATION AND RESEARCH**

How does evaluation compare with research? The words *evaluation* and *research* are sometimes used interchangeably. Although they hold similar characteristics, they are also different from each other.

Similarities: Both evaluation and research involve an investigation about something, and both are used to answer questions that are valuable to certain groups of people or organizational entities. They both employ data collection methods such as surveys, interviews, focus groups, observations, extant data reviews, or tests. The terms *evaluation* and *research* are often used interchangeably because similar data collection methods are used during evaluation and research. In both cases, the collected data are used as evidence to answer the proposed questions.

Differences: Evaluation is often conducted based on a client's request or approval, whereas most research, whether it is quantitative or qualitative research, is not client based. *Client-based* evaluations are usually reported directly to the clients and the client organizations rather than to the public, whereas research results are often presented in public venues such as research journals or conferences to allow the public as well as other researchers to have access to them.

Evaluation and research are also different in terms of their relationship with the context in which they are conducted. Evaluation activities are often *context-specific* (to investigate what happened in this particular context). In contrast, quantitative research activities are designed to be *context-free* (to ensure generalizability) and qualitative research activities can be characterized as *context-sensitive* (to allow transferability). Most evaluations produce context-specific evaluative conclusions about the quality, value, or significance of the subject being studied, whereas most research intends to produce generalizable or transferrable conclusions about the investigated research questions rather than drawing evaluative conclusions about a specific subject itself.

> **FIGURE 1 ● The Relationship Between Quantitative Research, Qualitative Research, and Evaluation**
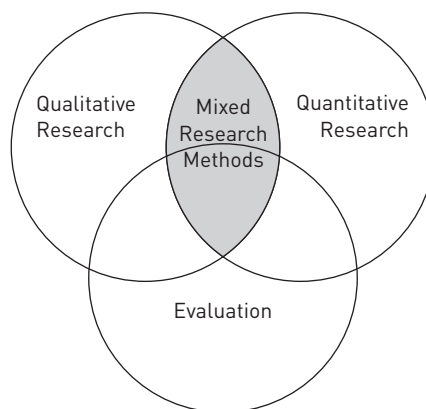
Relationship: Evaluation can be viewed as a subset of research, or vice versa. Evaluation and research can also be viewed as two ending points of a continuum or as overlapping with each other (Mathison, 2008). There is no one right way to define the relationship between evaluation and research. In this book, we will take the view that evaluation and research overlap. Figure 1 shows a Venn diagram illustrating the relationship between evaluation and research. Evaluation may use quantitative research, qualitative research, or mixed research methods, while possessing its own characteristics distinguishing itself from either type of research.

EVALUATION AND RESEARCH OVERLAP

Program Evaluation in the HPI Context

(Adapted from Chyung, 2015)[a]

Earlier, HPI was described as a transdisciplinary field of practice but not a discipline in itself. In contrast, evaluation (as well as research) is a transdisciplinary field of study and is recognized as a discipline (Scriven, 1991b). Evaluation, grown out of mere practice and now developed into its own discipline, crosses disciplinary boundaries aimed at helping to enhance performances of various professions. As illustrated in the examples of HPI application earlier, organizations in various industries would conduct evaluations to make improvements on their practice or products. Evaluation was destined to be included in the HPI practice because evaluation also seeks accountability and improvement of interventions.

EVALUATION AS A TRANSDISCIPLINARY FIELD

When HPI interventions are implemented to improve performance outcomes, the interventions often take the form of new or modified programs (including process-related programs), such as training programs, newly designed work processes,

incentive programs, or performance support systems. Therefore, evaluation conducted in the HPI context can be characterized as a type of *program evaluation*. Combining our understanding of HPI and evaluation and largely influenced by Scriven's (1991b) definition of evaluation, we can define **program evaluation** used in the HPI context as the systematic and systemic collection and analysis of information about the process and outcomes of a program in order to make improvements or judgments about the program's quality or value.

- The first part of the definition—"the systematic and systemic collection and analysis of information about the process and outcomes of a program"— describes the *means* used during program evaluation.

- The second part of the definition—"in order to make improvements or judgments about the program's quality or value"—describes the intended *outcomes* of conducting program evaluation.

This definition of program evaluation means that the overall evaluation plan and the data collection and analysis procedure should be executed in a step-by-step fashion (systematic), and the analysis of the information and determination of the program's value should be carried out by considering various interrelated factors and different parts of the organization and its outside community (systemic). In this book, the word *evaluation* refers to program evaluation conducted in the HPI context as defined earlier.

Evaluation Is Often Neglected

As explained earlier, the evaluation phase should interact with each phase of the human performance technology model through an exchange of information. However, despite the important role that evaluation plays in the integrated HPI process, organizations often think of evaluation as a post hoc activity or neglect it altogether. Organizations report that they do not always conduct evaluations after interventions have been implemented, and when they do, they do not produce comprehensive evaluations (Pulichino, 2007).

Let's take a look at the research conducted on training evaluations in particular, as organizations invest a tremendous amount of money on workforce learning and development. For example, in 2011, U.S. organizations spent nearly $156.2 billion on it (American Society for Training & Development, 2012), and in 2015, the average organization spent about 4.3% of payroll on direct learning expenditure, up from 4.0% in 2014 (American Society for Training & Development's, 2016). A popular approach to conducting training evaluations is to perform four levels of evaluations: reaction, learning, behavioral changes, and organizational results (Kirkpatrick, 1956, 1996a).

Research has shown that the frequency of evaluating behavioral changes and organizational outcomes has a significantly positive correlation with the trainees' knowledge and skill transfer (Saks & Burke, 2012). Research has also shown that organizations place high value on evaluations of behavioral changes and organizational

outcomes (American Society for Training & Development, 2009). However, there seems to be a large gap between such value estimation on evaluation and the actual conduct of evaluation in organizations. It has been reported that although organizations often conduct evaluations to investigate participants' reactions and learning, less than 50% of organizations conduct evaluations on participant's behavioral changes and organizational outcomes (Pulichino, 2007).

Barriers to conducting comprehensive evaluations include environmental factors, such as lack of resources (time and personnel) and lack of managerial support, and personal factors, such as lack of expertise in evaluation methodology (Kennedy, Chyung, Winiecki, & Brinkerhoff, 2014). Thus, you as an evaluator would need to overcome both types of barriers by improving your own evaluation skills while attempting to reduce environmental barriers to conducting evaluations. This book aims to help you develop evaluation-specific skill sets.

BARRIERS TO CONDUCTING COMPREHENSIVE EVALUATIONS

Different Evaluation Designs Used in Program Evaluation

Similar to conducting research, you as an evaluator need to possess a broad range of knowledge and skill sets for planning and executing your evaluation projects with different evaluation designs depending on the type of evaluation questions you need to answer. Among evaluation designs, experimental and descriptive evaluation designs are often used.

When you intend to show a cause-and-effect relationship between an intervention program and its outcomes, you may use an **experimental evaluation design**. You randomly select a sample of participants from the population and randomly assign them to different conditions (e.g., an intervention program used or no intervention program used) to see if the different conditions produce different outcomes. When it is not practical to use a random assignment method, your experimental study becomes a quasi-experiment.

DIFFERENT EVALUATION DESIGNS

The experimental design is considered the most rigorous approach when needing to show a cause-and-effect relationship between the intervention program and its outcomes, that is, if the program indeed caused the observed outcomes. However, to use an experimental design, evaluators need to set up ahead of time a group of people that use an intervention program and a nonintervention group who do not use it. This makes it difficult to use an experimental design in the HPI context because needs for evaluation often arise after intervention programs have been implemented (Wang, 2002).

Alternatively, you may use a **descriptive evaluation design** to study a case without manipulating any variables. The case can be a person, a group of people, or an organization, which is often purposely selected (a.k.a. purposive sampling). By generating descriptive information about the case that you are studying, you gain an in-depth understanding about the case. Although a descriptive evaluation design involves a case (or multiple cases), a case study does not always use a descriptive evaluation design; it may employ an experimental/quasi-experimental design as well. For example, you may conduct a program evaluation within a specific organization

(case), where you use a descriptive evaluation design to investigate environmental factors that influence employees' performance. Within the case study type evaluation, you may also use an experimental design with a random sample of employees to investigate the effectiveness of a proposed intervention on their performance outcomes.

Descriptive Case Study Type Evaluation Design

This book will explain to novice or advanced beginner level evaluators how to use a **descriptive case study type evaluation design** to produce a reasonably comprehensive evaluation. For example, consider the following program evaluations you can design with a descriptive case study type evaluation design.

Example 1: Several years ago, your company implemented a new on-site daycare program to some of the company's branches in the nation, and many employees have been using the program. Your boss asked you to conduct an evaluation of the new on-site daycare program to find out about its value. You identified the overall evaluation question as: *What values does the on-site daycare program provide to the employees?* To answer the evaluation question, you decided to investigate different aspects of the program by answering the following specific questions:

1. Quality of the daycare program curriculum—How well is the daycare program curriculum designed, compared to other programs' curricula?

2. Employee job satisfaction—How does the daycare program influence employees' job satisfaction?

3. Employee turnover—How does the daycare program affect employees' decision to stay or leave the company?

For Question 1, you review the daycare program curriculum and compare it to other programs' curricula. You also observe the daily activities for a few days. To investigate Questions 2 and 3, you use multiple data collection methods. You survey two groups of employees who need a daycare service: those who have participated in the on-site daycare program and those who used an off-site program. You randomly select several employees from the groups and interview them to learn more about benefits and challenges associated with using the on-site versus off-site daycare programs. You also analyze the turnover rates among the employees who have participated in the on-site daycare program and those who have not. You review notes from exit interviews conducted with those who voluntarily resigned or were terminated to see if they indicated daycare-related issues.

Example 2: Animal shelters often get help from volunteers but the volunteers' first-month dropout rate is usually high. You are conducting an evaluation of an animal shelter in your city to find out what can be done to reduce the volunteer dropout rate

and improve the quality of the volunteer program. You decide to focus on the following aspects of the volunteer program:

1. Training materials used for new volunteers—How well are the training materials designed?

2. Volunteers' goal achievement—Are volunteers achieving their goals? What are the barriers to achieving their goals?

3. Volunteers' burnout—Do volunteers feel burned out? What causes their burnout feelings?

4. Shelter visitors' satisfaction—How satisfied are visitors toward volunteers' work?

To investigate Questions 1 through 4, you collect data by reviewing existing training materials; surveying and interviewing volunteers, trainers, and visitors; and observing volunteers' training program and actual performance. You collect both quantitative and qualitative data and compare data obtained from multiple sources (volunteers, trainers, and visitors), combine results obtained from multiple aspects of the program in order to draw conclusions and provide recommendations for improvement.

As illustrated in these examples, this book explains how you can conduct a program evaluation by using a descriptive case study type evaluation design. As a novice or advanced beginner level evaluator, you want to start with a small evaluation project for an internal or external client, rather than conducting a high-stakes evaluation. You should continue to develop your evaluation competency level up to competent, proficient, and expert levels by acquiring more knowledge and skills to conduct evaluations using other types of evaluation designs. In addition, while navigating through organizational politics, you need to develop skills for effectively communicating with stakeholders to discover and align their needs with intended and unintended outcomes of the program.

Frameworks for Conducting Evaluations in the HPI Context

Recall that we defined program evaluation conducted in the HPI context as:

the systematic and systemic collection and analysis of information about the process and outcomes of a program in order to make improvements or judgments about the program's quality or value.

Because various types of intervention programs are used for workforce development and performance improvement, HPI practitioners need expertise on conducting evaluations of various types of programs rather than just training programs. In fact, research has shown that only 10.5% of performance improvement interventions are knowledge-improvement programs such as training and education programs

(Gilbert, 1997). Katzell-based Kirkpatrick's four-level training evaluation framework (Kirkpatrick, 1956; Smith, 2008) is designed to help you conduct training evaluation. However, it is not a comprehensive model for conducting evaluations of the remaining 89.5% of nontraining programs. Thus, you need additional frameworks that can be applied to evaluating either training or nontraining interventions.

MULTIPLE EVALUATION FRAMEWORKS AS WELL AS PROFESSIONAL STANDARDS AND PRINCIPLES

There are many evaluation frameworks and approaches to conducting program evaluations. One source lists 26 approaches to evaluation (Stufflebeam & Shinkfield, 2007). There is no one particular evaluation approach that is applicable to all types of evaluations. It is impossible to cover all of the available evaluation frameworks in a book. Therefore, in this book, you will be introduced to the following frameworks (in alphabetic order) that were selected to help you understand the fundamental principles and procedures associated with conducting a program evaluation in the HPI context in a systematic and systemic way (a summary of each is described in Appendix A):

1. Behavior engineering model (Gilbert, 2007)

2. Four levels of training evaluation (Katzell, n.d., as cited in Kirkpatrick, 1956; Kirkpatrick, 1996a)

3. Key Evaluation Checklist (Scriven, 2013)

4. Program logic model guidelines (W. K. Kellogg Foundation, 2004)

5. Success Case Method and training impact model (Brinkerhoff, 2006)

6. Utilization-focused evaluation (M. Q. Patton, 2012)

You will also apply evaluation-specific professional principles and standards, such as the following:

- American Evaluation Association (AEA)'s guiding principles for evaluators (www.eval.org)

- The Joint Committee on Standards for Educational Evaluation (JCSEE)'s program evaluation standards (www.jcsee.org)

Based on these evaluation frameworks and professional standards and principles, we recognize multiple ingredients necessary for conducting a program evaluation (Figure 2).

The 10-Step Evaluation Procedure

The multiple ingredients shown in Figure 2 are incorporated into a 10-step evaluation procedure, which is outlined in Table 1. Michael Scriven's work, especially his **Key Evaluation Checklist** (Scriven, 2013), was instrumental in developing the systematic 10-step evaluation procedure. Many of the main concepts and steps used during the 10-step evaluation procedure, such as evaluand, three types of impactees,

FIGURE 2 ● Ingredients Needed to Conduct a Comprehensive Program Evaluation

Evaluation Ethics

Meta-Evaluations

Rubrics for Quantitative and Qualitative Data

Synthesis of Dimensional Results

Evidence-Based Practices

Instruments and Data Collection

Importance Weighting Among Dimensions

Feasibility and Risks

Dimensions (Including Kirkpatrick's Four Levels)

Stakeholder Analysis

Program Logic Model or Training Impact Model

Formative or Summative

Goal-Based or Goal-Free

Evaluand and Stakeholders

To determine or improve the quality or value of a program

goal-free evaluation, dimensions, importance weighting, synthesis, meta-evaluation, and more, are based on Scriven's work.

The 10 steps are described as *micro*-level steps. You also perform three *macro*-level tasks such as assessments of feasibility, assessments of risk factors, and meta-evaluations. Each of the macro-level tasks is not just a single step to complete but rather an ongoing application throughout the project. As an analogy, the micro-level systematic 10 steps should be immersed in systemic application of the macro-level tasks, as illustrated in Figure 3 (Barkin, Chyung, & Lemke, 2017).

THREE PHASES OF EVALUATION AND THREE DELIVERABLES

The 10 steps are divided into three phases—identification, planning, and implementation—producing three deliverables (Figure 4). During the **identification phase**, you communicate with your client to identify and clearly understand the program to be evaluated (a.k.a. an evaluand) (Step 1), its stakeholders (Step 2), and the purpose of the evaluation (Step 3). You will likely perform these first three steps simultaneously. You also assess feasibility and risk factors for the evaluation project. If feasible, you prepare the first deliverable to be submitted to your client, which likely takes the form of an agreement or contract regarding the evaluation to be performed. This agreement is called a statement of work (SOW). However, your SOW will not contain detailed information about exactly how you will perform the evaluation. In some cases, where you conduct an evaluation as an internal evaluator or part of your job responsibility, you may proceed to the planning phase without producing a SOW.

During the **planning phase** of your evaluation project, you learn more about the inner workings of the program with a program logic model (Step 4), identify aspects (dimensions) of the program to be investigated (Step 5), and determine the data collection methods you will use (Step 6). You continue to monitor the program feasibility and risk factors. You also start conducting formative meta-evaluations to ensure that you are developing an appropriate evaluation plan. At the end of the planning phase, you are ready to submit your second deliverable to your client—a comprehensive evaluation proposal.

The identification and planning phases of evaluation can also serve as an evaluability assessment stage (Trevisan & Walser, 2015), during which you determine whether or not a program is ready for evaluation, how interested the stakeholders are in conducting the evaluation, and how the stakeholders will use the evaluation results. When the program itself has not quite matured enough to be evaluated for its outcomes, pursuit of an outcome-based evaluation will only yield unreliable results, often showing nonsignificant improvement. In such cases, having an evaluability assessment stage will help you determine whether or not to continue with the evaluation project or recommend postponement until the program develops more and starts producing reliable outcomes.

If the program is determined to be ready for evaluation and your client approves your evaluation proposal, you move on to the **implementation phase**. You develop data collection instruments (Step 7), collect data with the instruments (Step 8), analyze data with rubrics (Step 9), and draw conclusions (Step 10). You perform formative meta-evaluations throughout these steps. At the end of this last phase, you prepare your third deliverable—an evaluation report—and you conduct a summative meta-evaluation. Then, you submit your final report to your client and other stakeholders and assist them in gaining utility and applicability from the findings.

Detailed explanations on how to follow the 10-step evaluation procedure are provided in the rest of this book.

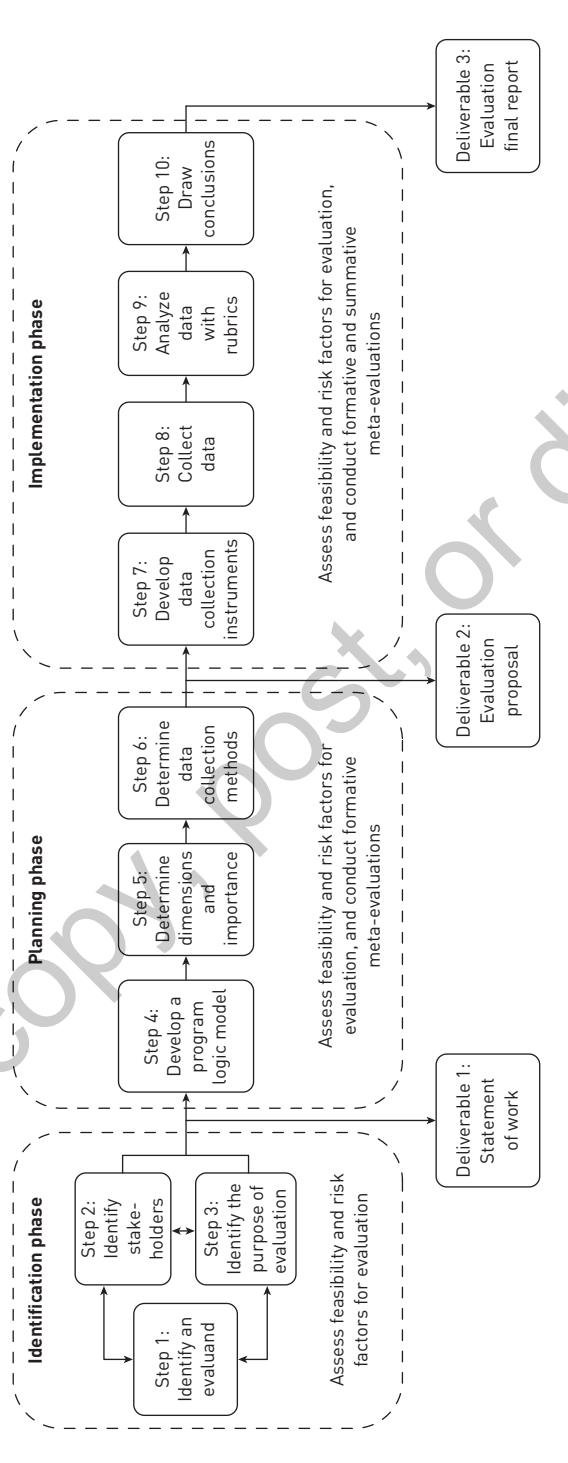| TABLE 1 ● The 10-Step Evaluation Procedure				
Macro-level	Micro-level	Main Task	During the Step	Deliverable
Assess feasibility, assess risk factors, and conduct meta-evaluations.	1	Identify the program to be evaluated.	Meet with the client to learn what needs to be evaluated.	*Identification phase:* Submit a statement of work (SOW) to the stakeholders and continue with evaluation planning or decline the evaluation request.
	2	Identify stakeholders of the program and their needs.	Understand the program stakeholders' needs.	
	3	Identify the purpose of evaluation based on how the evaluation findings will be used.	Steps 1, 2, and 3 may run in parallel to finalize the purpose of evaluation based on stakeholders' intention to use the findings of the evaluation.	
	4	Develop or review a program logic model for the program.	Involve the client and other stakeholders.	*Planning phase:* Submit an evaluation proposal to the stakeholders and continue with the implementation phase, or forgo/delay it.
	5	Determine dimensions and importance weighting.	Align dimensions with the stakeholders' needs and their intention to use evaluation findings.	
	6	Determine data collection methods.	Plan to use direct measures and multiple data sets whenever possible.	
	7	Develop data collection instruments.	Be sure to develop valid and reliable instruments. Obtain approval from the stakeholders.	*Implementation phase:* Submit an evaluation final report to the stakeholders and assist their use of the evaluation findings.
	8	Collect data.	Maintain confidentiality of data and handle human subjects and data ethically.	
	9	Analyze data with rubrics.	Triangulate multiple data sources and check with the stakeholders regarding appropriateness of rubrics.	
	10	Draw conclusions.	Organize findings based on the stakeholders' needs and their intention to use evaluation findings.	

FIGURE 3 ● Micro-Level Systematic 10 Steps Immersed in Systemic Application of Macro-Level Tasks[b]

A 10-Step Evaluation Procedure

Identification phase:
1. Identify the program to be evaluated
2. Identify stakeholders of the program and their needs
3. Identify the purpose of evaluation based on how the evaluation findings will be used

Deliverable: Statement of Work

Planning phase:
4. Develop or review a logic model for the program
5. Determine dimensions and importance weighting
6. Determine data collection methods

Deliverable: Evaluation Proposal

Implementation phase:
7. Develop data collection instruments
8. Collect data
9. Analyze data with rubrics
10. Draw conclusions

Deliverable: Evaluation Final Report

Feasibility

Risk factors

Risk factors

Risk factors

Meta-evaluation

Risk factors

Feasibility

Meta-evaluation

---

[b] The draft version of this book was the source of the figure cited in the following article: Barkin, J. R., Chyung, S. Y., & Lemke, M. (2017). Following a 10-step procedure to evaluate the administrative services qualification card program. *Performance Improvement, 56*(8), 6–15. doi:10.1002/pfi.21717

FIGURE 4 ● **Three Phases of Evaluation With 10 Micro Steps and Three Macro Tasks to Produce Three Deliverables**

Identification phase

Step 1: Identify an evaluand

Step 2: Identify stake-holders

Step 3: Identify the purpose of evaluation

Assess feasibility and risk factors for evaluation

Deliverable 1: Statement of work

Planning phase

Step 4: Develop a program logic model

Step 5: Determine dimensions and importance

Step 6: Determine data collection methods

Assess feasibility and risk factors for evaluation, and conduct formative meta-evaluations

Deliverable 2: Evaluation proposal

Implementation phase

Step 7: Develop data collection instruments

Step 8: Collect data

Step 9: Analyze data with rubrics

Step 10: Draw conclusions

Assess feasibility and risk factors for evaluation, and conduct formative and summative meta-evaluations

Deliverable 3: Evaluation final report

Chapter Summary

- Evaluation, including human performance improvement (HPI), is a transdisciplinary field of practice.

- Evaluation is one of the essential components in the HPI process, and it should interact with each phase of the HPI process.

- Evaluation involves value assignment; thus, evaluation is not the same as measurement or assessment, which does not involve a value judgment.

- Evaluation and research have similarities and differences; they can be described as one being a subset of another, two end points of a continuum, or overlapping with each other.

- Evaluation conducted in the HPI context can be characterized as a type of program evaluation. Program evaluation conducted in the HPI context is defined as the systematic and systemic collection and analysis of information about the process and outcomes of a program in order to make improvements or judgments about the program's quality or value.

- Evaluation can use different evaluation designs, including experimental, quasi-experimental, and descriptive designs.

- Evaluation and research can use various data collection methods, such as self-administered surveys, interviews, focus groups, observations, extant data reviews, and tests.

- Evaluators can conduct program evaluations in the HPI context by using various frameworks derived from the HPI and evaluation fields. In doing so, they should apply evaluation-specific professional standards and principles.

- This book is written for novice or advanced beginner level evaluators to conduct evaluations of instructional or noninstructional programs by using a descriptive case study type evaluation design. To develop evaluation skills to competent, proficient, and expert levels, evaluators should learn to conduct program evaluations with other types of evaluation designs.

- This book explains a 10-step procedure of conducting program evaluations in the HPI context; the 10 steps are divided into three phases—identification, planning, and implementation—producing three deliverables—a statement of work, an evaluation proposal, and an evaluation final report.

Chapter Discussion

1. How good is this apple?

Every day, we perform various evaluations, small or large, informal or formal. For example, you go to a grocery store and may have to decide which apples to buy—your decision depends on whether you are

making an apple pie or you want an apple just for munching. Similarly, you may be looking around to buy a car, and your decision will depend on whether it is going to be used as a commuter car for yourself or a family car for your six-member family.

Regardless of what you are evaluating, you go through a similar evaluation process. Let's have a small fun activity with your friends or family members to document the evaluation process.

1. Describe the thing that you want to evaluate: e.g.,

   A bag of apples (based on a sample piece)

2. Describe who will use it: e.g.,

   My family, including two adults and two adolescent children

3. Describe where and why they will use it (context and purpose): e.g.,

   Believing in the saying, "An apple a day keeps the doctor away," family members eat an apple a day as a snack.

4. As shown in the example in Table 2, discuss with the users, and do the following:

   a. List three or four dimensions (aspects) that its users will use to judge its quality.

   b. Rank-order the dimensions in terms of its importance to the users.

   c. Determine a rubric you will use to judge each dimension's quality.

   d. Determine a score for each dimension by measuring it against the rubric.

   e. Determine the overall quality based on the combined results against a final rubric.

   f. Make a data-driven decision.

5. Discuss how this process can be applied to an evaluation of a performance improvement intervention in the HPI context.

2. What individual effort and organizational support are needed to successfully conduct program evaluations in organizations?

Based on your observation at the organization where you currently work or previously worked, how often are performance improvement interventions (programs) evaluated after they are implemented? What are the drivers for and barriers to evaluating performance improvement programs in a systematic and systemic way?

Using Table 3, identify things that you as an individual practitioner and the organization can do to facilitate and support the systematic and systemic evaluation of performance improvement programs.

TABLE 2 ● How Good Is This Apple?				
Dimensions	**Importance Weighting**	**Dimensional Rubrics**	**Dimensional Finding**	**Overall Quality**
List dimensions that its users will use to judge its quality.	Rank-order dimensions in terms of its importance to users: e.g., 1. Important 2. Very important 3. Critical	Develop dimensional rating rubrics with 3-4 levels of descriptions: e.g., 1. Nah 2. OK 3. Awesome	Measure (try the sample piece) and rate each dimension against the dimensional rubric.	Factoring the importance weighting into the dimensional results, develop a final rubric and determine the overall quality. Then, make a final decision.
A. Taste	3. Critical	1. Nah: Bitter tart 2. Not bad: Sweet and a little bit tart 3. Awesome: Really sweet!	It is rated as *Not bad* because it is mostly sweet although it has a little bit of tart taste.	Final rubric: 1. Poor (I would not buy it): If at least one dimension = Nah 2. It's OK (I may buy it if there are no better options): If Critical dimension = Not bad, and other dimensions = Not bad or Awesome 3. Excellent (I will certainly buy it again): If Critical dimension = Awesome, and other dimensions = Not bad or Awesome
B. Size	1. Important	1. Nah: Half of my fist size 2. Not bad: Two thirds of my fist size 3. Awesome: As large as my fist	It is rated as *Awesome* because it is about my fist size.	
C. Texture	2. Very important	1. Nah: Mushy 2. Not bad: Firm 3. Awesome: Crunchy	It is rated as *Not bad* because it is firm enough.	**Overall quality: "It's OK."** **Final decision: I will buy it because there are no better options.**

Adapted from Davidson, 2005, pp. 151–187; Preskill & Russ-Eft, 2005, p. 20.

TABLE 3 ● Identify Individual Effort and Organizational Support Needed for Conducting Systematic and Systemic Evaluation	
Individual effort	Organizational support
•	•
•	•
•	•
•	•