Chapter 23
Demystifying Program Evaluation for Surgical Education

Alexis Battista, Michelle Yoon, E. Matthew Ritter, and Debra Nestel

Overview In this chapter, we define program evaluation, address its role in evaluating surgical education programs, describe important early steps surgical program evaluators can take to improve the usefulness of program evaluation, discuss common challenges, and offer solutions evaluators can use to overcome these challenges. The chapter is intended for those who are engaging or considering engaging in program evaluation for the first time or are doing so with limited support from a formal program evaluator. Additionally, we have included resources and examples to provide guidance beyond the scope of this chapter.

23.1 Introduction

There are times when policymakers, accreditation organizations, university or hospital leadership, and program and clerkship leaders will ask questions about the effectiveness of their surgical education programs and interventions. Evaluating surgical programs can answer questions such as:

A. Battista (✉)
Graduate Programs in Health Professions Education, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

The Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA
e-mail: alexis.battista.ctr@usuhs.edu

M. Yoon
University of Colorado School of Medicine, Aurora, CO, USA

E. Matthew Ritter
USU/Walter Reed Department of Surgery, Bethesda, MD, USA

D. Nestel
Faculty of Medicine, Nursing & Health Sciences, Monash Institute for Health and Clinical Education, Monash University, Clayton, VIC, Australia

Faculty of Medicine Dentistry & Health Sciences, Department of Surgery, Melbourne Medical School, University of Melbourne, Parkville, Australia
e-mail: dnestel@unimelb.edu.au; debra.nestel@monash.edu

D. Nestel et al. (eds.), *Advancing Surgical Education*, Innovation and Change in Professional Education 17, https://doi.org/10.1007/978-981-13-3128-2_23

- What is the nature and scope of a surgical education program problem? Whom does it impact, how many are affected, and how does the problem affect them?
- What are the possible intervention options that are likely to ameliorate a defined problem in surgical education?
- Is the surgical residency error reduction program attaining the desired goals and benefits?
- Is a new surgical fellowship training program being implemented well?
- Is the quality improvement intervention program changing surgical outcomes?
- Is the cost of the simulation-based skills training program reasonable when compared to its effectiveness and benefits?

Seeking and providing these answers is increasingly the work of surgical program leaders, such as clerkship, residency and fellowship directors who may be asked to study, appraise, and improve surgical education programs. The purpose of this chapter is to provide a practical overview of program evaluation (PE) and how to get started. This chapter does not present a comprehensive discussion on PE but instead provides information and guidance about how to get started while pointing to additional resources that can be used to enhance evaluation efforts in the future.

23.2 Defining Program Evaluation

PE is the use of research methods to systematically investigate the effectiveness of social and educational programs, including surgical education and interventions, to guide future efforts, and to change or improve a program [1, 2]. *Educational evaluation* is the process of defining, gathering, analyzing, and disseminating information to guide decisions about an educational program [3, 4]. The target of an evaluation may be any organized educational program, including:

- A curriculum
- A course
- A specific instructional approach (e.g., simulation-based learning, journal club)
- Policies and guidelines
- Specific services that are part of the educational experience

PE seeks to address questions of need for, quality of, processes of, or the impact of an educational program in the context of continuous quality improvement and decision-making [5, 6]. Evaluations can focus on whether an educational program is working as intended or if there are unintended consequences [7]. In the case of surgical education, PE may be conducted on a surgery clerkship rotation, residency program, or fellowship training program. It may entail evaluating the overall program outcomes (e.g., data about satisfaction at the end of training, exam performance, job placement of graduates), as well as examining granular-level pieces such as individual teaching sessions or the teaching and learning structure or environment.

Evaluation may occur on a program that is currently in progress (formative evaluation) or on completion (summative evaluation). Both types of evaluations can help program stakeholders make decisions about what should be kept or changed in a program or even determine whether a program should continue [1, 2]. In surgical education, PE usually provides information about the effectiveness of educational training programs with the comprehensive purpose of optimizing healthcare outcomes and quality.

23.2.1 Why Conduct a Program Evaluation: What's in It for You?

PE can provide information that can create value at the institutional level as well as at the accreditation level. For example, assessment data, such as qualifying professional examination pass rates, can yield information about program outcomes, while PE data collected over time can provide insights into job placement and long-term program impact [7–9]. Additionally, PE findings can be used to inform or provide feedback to faculty who provide instruction in a program, which in turn, may be useful for faculty career development or promotion. Findings from a PE may also provide feedback to administrators, support staff, and others who are instrumental in maintaining a program's structure and logistic operations (e.g., library, technology, assessment). Furthermore, data generated from PE can be used to inform the design and implementation of surgical education programs which should be viewed as a cycle of *designing*, *implementing*, *evaluating*, *and revising*, rather than a static state [10]. PE supports this cycle because the results can inform continuous quality improvement in the instructional and curriculum design process by providing more precise information about what works, what doesn't work, and what could be altered [7, 10]. In turn, the results of PE can also support the efforts of accreditation and reaccreditation because accreditation bodies (e.g., American College of Surgeons) may require program managers to report and share student outcomes as well as demonstrate that the program engages in regular evaluation efforts.

23.2.2 Resources and Guidelines in Surgical Education

Highlighting the growing importance of PE in surgical education, many surgical associations offer guidance, standards, and resources for PE including the American College of Graduate Medical Education (ACGME), the Royal Australasian College of Surgeons (RACS), and the Royal College of Surgeons (RCS). It is worth reviewing and considering the guidelines from accrediting groups because they often provide specific criteria that can influence or guide your PE. Furthermore, although surgical department faculty often conduct PE, it is also important to be aware of the

role the professional evaluators may play. Professional evaluators are individuals with diverse training and professional experience in the practices of evaluation. Professional evaluators can play an important role in improving the design or implementation of an evaluation, particularly if your funding agency requires it, if the PE you are planning is highly complex, or if you determine that including a professional evaluator can add credibility to the PE process or findings. For additional information about professional evaluators and their role, the American Evaluation Association (AEA) is a valuable resource.

23.3 Getting Started: Key Stages of Conducting a Program Evaluation

Although there are many steps included in any PE, one of the most challenging is deciding where and how to start. Importantly, a key aspect of getting started is keeping in mind that there is no "one size fits all" in PE [1]. For PE to be successful, it must be tailored to the unique needs of the organization. The most successful evaluations are ones that provide useful and credible information that support decision-making [1]. In this section, we discuss key steps of PE to support efforts in getting started and ensuring your evaluation is tailored to your organization's unique needs. These steps include identifying and involving stakeholders, developing a logic model, focusing your evaluation, and selecting an evaluation model.

23.3.1 Identifying and Involving Stakeholders

Identifying stakeholders early in an evaluation is a key step to ensuring the evaluation will yield useful information. Stakeholders include both people and entities, who are or may be affected by the program under evaluation [1]. Stakeholder identification sets the stage for the entire PE process and can help generate useful evaluation questions and help identify human and financial resources to conduct the evaluation and targets for dissemination of the findings.

On first pass, surgical educators new to PE may not appreciate the scope and importance of identifying and involving stakeholders. At the operational level, clerkship and program directors might quickly identify the need to respond to demands of the Chairman, regulatory bodies (e.g., ACGME, RCS, RACS), or to highlight successes to aid in the recruitment of future trainees; however, the scope and importance of stakeholder involvement extend well beyond this initial focus. Therefore, it is important that the evaluator identify a broad list and make an informed decision about each stakeholder's level of participation.

In keeping with the practical focus of this chapter, we highlight Green's (2005) conceptual framework for identifying potential stakeholders for PE [11]. Green's approach involves identifying stakeholders in one of four groups: those who have authority over the program, those responsible for the delivery of the program,

Authority	***Delivery***
➢ Department chairman ➢ Designated institutional official ➢ Dean's office ➢ University leadership ➢ Hospital leadership ➢ University / Hospital board of directors ➢ Philanthropic donors ➢ Accrediting groups ➢ Government or funding agency	➢ Program / Clerkship directors ➢ Program / Clerkship coordinators ➢ Core / Affiliated faculty ➢ GME Office staff ➢ Simulation center faculty / Staff ➢ Educators from disciplines other than surgery (e.g. human factors, sociologists etc.)
Program Beneficiaries	***Those Potentially Disadvantaged by the Program***
➢ Fellows / Residents / Students ➢ Spouses / Significant others ➢ Patients /Patent's family	➢ Programs with clinical overlap (e.g., effect of fellowships on residencies) ➢ Allied health training programs / trainees ➢ Private / Group practices without trainees

Fig. 23.1 Examples of categorizing stakeholders drawn from Green's [11] approach to identifying potential stakeholders and their roles

intended beneficiaries of the program, and lastly and most often overlooked, those who may be disadvantaged by the program. Drawing from this model, Fig. 23.1 gives examples from surgical education for each of these categories of stakeholders.

Once identified, the role of each stakeholder in the evaluation process needs to be defined. One way to approach this is to assign a primary role to each identified stakeholder. Categorical levels of participation ranging from least involved to most involved could include awareness, policy and guidelines, input and reaction, and operational decision-making. While none of these categories need to be mutually exclusive or absolute, this organization helps the evaluator systematically consider how stakeholders may influence the evaluation. There is no correct or incorrect way to do this, but the approach should be adapted to each situation. A thoughtful approach to determining the involvement of key stakeholders will allow the PE to have the most meaningful impact.

23.3.2 Developing a Logic Model

Simply put, it will be difficult to evaluate how well a program is doing or working if stakeholders don't have an explicit understanding of what the program is supposed to be doing and how it is supposed to work (also called a program theory).

Therefore, when designing a PE, it can be useful to develop a *logic model*. A logic model is a graphic representation that helps visually represent the connections between the "if-then" causal relationships of the program activities (e.g., inputs such as educator time and teaching materials and outputs such as short- or long-term goals) [1, 12, 13]. By making these causal relationships explicit, stakeholders can make better judgments about programs processes or efficacy, which, in turn, can help improve the usefulness of the PE [1]. In the event that there isn't a clear understanding of the programs' theory, the focus of the PE may emphasize developing a logic model. Notably, logic models are also increasingly required in grant programs and global surgery projects. For more comprehensive details about how to develop a logic model, we refer readers to McLaughlin and Jordan (1999), Shakman and Rodriquez (2015), and Lawton and colleagues (2014), to name a few [12–14].

23.3.3 Focusing Your Evaluation

Once stakeholders are identified, their roles defined, and a logical model is developed, it is helpful to employ a systematic approach to further focus and plan the PE. Although there are several ways to organize and focus a PE, we highlight one framework that has been adapted from a 10-step approach presented in the American College of Surgeons "Surgeons as Educators" course. Figure 23.2 demonstrates examples of how to apply these steps.

It is important to remember that the steps outlined above help generate a *focused and comprehensive plan* for PE – a process more akin to a marathon than a sprint. For example, it is often best to generate the right questions and work to answer them, rather than only asking questions that you can answer with the data you have. Avoiding this common pitfall will result in a substantial improvement in both the evaluation and, importantly, the program itself. Once your initial plan is in place through application of the 10-step model or another framework, selecting an evaluation model that is grounded in sound measurement theory can help move the process along.

23.3.4 Selecting an Evaluation Model

Conducting an evaluation is a complex task, particularly if you are new to conducting evaluations, have limitations to paying for, or accessing, external resources (e.g., program evaluation professional), or, like many clinicians, are juggling evaluation efforts with your teaching and clinical responsibilities. Using and adapting an existing evaluation model can help demystify the process of conducting a PE by offering structure and support while guiding decision-making processes and methodological choices [1]. Using a model also helps assure that important steps and information are not overlooked or missed [1]. There are numerous models and approaches to PE.

	Step	Example
1	Identify key stakeholders	See previous section and Figure 23.1
2	Define the evaluation purpose(s) and how the results will be used	• Identify ability of the current program to meet upcoming changes in regulatory requirements. • Identify targets for cost savings without impacting quality of education.
3	Identify what should be evaluated and generate evaluation questions. (see following section on selecting an evaluation model)	• What is the percentage of high performing students who match in surgical residencies? • How do fellowship directors perceive the incoming performance of residency graduates?
4	Inventory what performance evaluation data are currently being collected and by whom	• Standardized test scores, Patient satisfaction questionnaires, Individualized quality data, Centrally administered surveys
5	Match data being collected with evaluation questions and determine need for new/additional data collection	• An evaluation question of first time Board pass rates may be answered by obtaining existing data. • A question about learner perception of faculty teaching effectiveness may require development of an assessment survey.
6	Develop timeline and responsibility for data collection	• Students must turn in completed clerkship patient logs before taking the subject exam. • Faculty must complete trainee assessments within two weeks of the end of the rotation.
7	Specify the analysis procedure to be used for each type of data and question.	• Effectiveness of a clinical rotation could be judged by the number of defined category operative cases or by themes in the narrative comments of residents on the post rotation survey.
8	Specify criteria to be used to make judgments (i.e. define “success”)	• Define an “acceptable” and “goal” for each evaluation question. • This could be achieved by using percentile ranks (e.g., above 50th percentile nationally for operative trauma volume) or by quantifying frequency of categorical themes based on narrative comments (e.g., positive comments regarding faculty support).
9	Determine which evaluation questions can be answered within your timeline, budget, and resources & identify what’s needed to answer all major evaluation questions	• A program director without protected time or administrative support can likely answer questions relating to case volume and first time board pass rates, but will not have the time to do in-depth analysis on faculty teaching effectiveness. • Stakeholders must help provide resources to answer the questions they helped generate.
10	Communicate results and follow up with key stakeholders	Socialize the results of the program evaluation. Meetings with underperforming faculty, sites, or affiliated programs will help ensure expectations are communicated and all factors are considered. Making leadership aware of both successes and challenges can facilitate the time and resources needed for improvement. Trend important results over time.

Fig. 23.2 Worked example of steps for planning a surgical program evaluation. (Adapted from the American College of Surgeons as Educators Workshop)

In medical and surgical education, some common approaches include *Kirkpatrick's Hierarchy* [15]; Patton's *Utilization-Focused Evaluation* (also called "Use-Based" Evaluation) [16]; Stufflebeam's *Context*, *Input*, *Process*, *and Product (CIPP)* model [17]; and *Outcomes-Based Evaluation* (also see Chap. 34). Table 23.1 provides a summary of these models, their key features, and their advantages and limitations.

23.4 Examples of Program Evaluation in Surgical Education

Table 23.1 demonstrates that the goals and approaches to PE vary widely. In a US-based study, Torbeck et al. (2014) describe an approach to evaluating a surgical residency program using an outcomes assessment system as a component of PE [18]. They use diverse data associated with one key stakeholder – the surgical resident – to track before, during, and after the surgical residency program. Collecting data over an extended time frame and for different cohorts helped identify features of the program while also informing decisions about what program features to maintain and what to strengthen.

In a second US-based study, Gomez et al. (2014) report an evaluation of an international medical student surgery-oriented program [8]. Like Torbeck et al. [18], the PE described by Gomez and colleagues has a strong outcomes-based focus drawing data from just one stakeholder – the students enrolled in the program. The findings of the evaluation are discussed in the light of broader macro level issues such as the forecasted decline in international medical graduates applying for residencies nationally [8].

In a third example, Yu et al. (2016) report a PE designed to develop competent cataract surgeons in China [9]. The program was comprised of two phases and focused on one procedure – phacoemulsification. Surgical trainee performance data and complication rates of patients were monitored in each phase and 2 years after attending the program. Although the complication rates fluctuated, performance improved across the program. The improvements were attributed to the programs combined features – wet lab exposure, deliberate practice with patients, and regular formative feedback using the performance measurement tool.

These examples demonstrate diverse ways in which these programs tailored their evaluations by including specific stakeholders and using different types of PE approaches (e.g., learning and career outcomes, patient complication rates) which helped them demonstrate the value of their programs.

Table 23.1 Common evaluation models used in health professions PE, key attributes, and advantages and limitations

Evaluation model	Attributes	Advantages	Limitations
Kirkpatrick's hierarchy or four-level evaluation	Widely recognized in health professions PE	Practical way to examine different levels of learner outcomes	Does not account for factors that may influence learning (e.g., motivation)
	Places an emphasis on program outcomes	Outcomes measures are commonly accepted by stakeholders	Does not describe how learning outcomes are supported
	Often combined with other evaluation approaches		
Use-based evaluations	Emphasis is on the needs and issues of the intended users, such as students, faculty, or other key program stakeholders	Focusing on intended users increases the chances that the findings will be useful and applicable	Placing the focus on the intended users can lead to other important viewpoints being overlooked
	Can be employed for formative and summative program evaluation	Can be used for a variety of program evaluation questions	
Context, input, process, and product (CIPP)	Links evaluation with decision-making and asks	Very systematic	May be very strict
	What needs to be done and were important needs addressed?	Comprehensive	Tends to be a "top-down" approach
	How should it be done?	Focuses on decision-making	May require more time to complete
	Is it being done?		
	Is the program succeeding?		
	Can be employed for formative and summative program evaluation		
Outcomes-based evaluations	Sometimes referred to as impact evaluation which focuses on exploring selected effects of a program	Can help identify immediate, short-, intermediate, or long-term program impacts	May not associate links between process and outcomes and so may not identify why a program is working (or not)
	Usually focuses on participants of the program although secondary or indirect audiences may also be considered (similar to the level 4 of Kirkpatrick above)	May uncover unintended outcomes or consequences	

23.5 Overcoming Challenges and Tensions in Program Evaluation

In the previous sections, we have highlighted and discussed key stages and processes associated with conducting PE. Although these steps suggest a linear and stepwise approach, the actual practice of PE can present a number of challenges. Some common challenges include:

- Staying the course when conducting a complex evaluation
- Considering how and where to report your PE
- Differences between PE and assessment
- Differences between PE and research

23.5.1 Staying the Course When Conducting a Complex Evaluation

The conduct of a PE can be overwhelming, particularly if you are new to PE or if you are juggling evaluation, teaching, and clinical duties. Additionally, some stages of PE may be more complex or take longer than others, or you may become aware of new evaluation questions and needs as your evaluation progresses. Some potential strategies for managing these challenges include:

- Use of a PE model helps guide and direct decision-making and helps minimize missed steps or stages of a PE.
- Break the PE into smaller, manageable tasks. This can offer some satisfaction that the PE is progressing while also providing you evidence of progress that can be shared with stakeholders when they request an update [1].
- When meeting with stakeholders, ask who might be available to help with the workload. As the evaluation progresses, continue to look out for additional supporters and resources.
- Develop and maintain a list of possible future evaluation questions, resources, and data sources. You may not pursue every new avenue, but keeping track of them can help you stay focused on your current evaluation plan.

23.5.2 Planning to Report and Disseminate Findings

It is important to think about reporting and dissemination at the outset of the PE planning process. Although most PE outcomes are reported textually, they may be accompanied by oral presentations before or after the release of a report. The sequence will vary with the purpose of the evaluation and stakeholder preferences. Additionally, there may be interim reports requested that may have different levels

I. Executive Summary
II. Introduction to the report
  a. Purpose of the evaluation
  b. Audiences for the evaluation report
  c. Limitations of the evaluation
  d. Overview of report contents
III. Focus of the evaluation
  a. Description of the evaluation object
  b. Evaluative questions used to focus the study
  c. Information needed to complete the evaluation
IV. Brief overview of evaluation plan and procedures
V. Presentation of evaluation results
  a. Summary of evaluation findings
  b. Interpretation of evaluation findings
VI. Conclusions and recommendations
  a. Criteria and standards used to judge evaluation object
  b. Judgements about evaluation object (strengths and weaknesses)
  c. Recommendations
VII. Minority reports or rejoinders (if any)
VIII. Appendices
  a. Description of evaluation plan/design, instruments, and data analysis and interpretation
  b. Detailed tabulations or analyses of quantitative data, and transcripts or summaries of qualitative data
  c. Other information, as necessary

Fig. 23.3 Structure of a PE Report. (Adapted from Fitzpatrick et al. [19])

of formality. Interim reporting is beneficial because it provides evaluators and stakeholders with an opportunity to engage in a dialogue about the progress of the evaluation. Interim reporting can also alert the evaluator to issues that might be important to include in the final report, which can potentially save time and increase the credibility of the findings. The audiences of the PE report may also vary, so tailoring of data may also be required in terms of the degree of detail, language style, and format. Additionally, when reporting, the evaluator must consider the ethics of PE practice to ensure accuracy, balance, and fairness [19]. Figure 23.3 highlights Fitzpatrick et al.'s (2011) structure for a written evaluation report [19].

23.5.3 Evaluation or Research?

In our work as program evaluators, we have often been asked, "Isn't this research? How does PE differ from research?" Confusion can occur because PE and research use similar methods. However, key differences include the purpose of the activity and the intended audience. (See Chap. 34 for further discussion of the differences.) When an evaluation reveals really interesting findings that evaluators think may be

of benefit to others, they may want to share this with a wider audience. This can present challenges because, unlike research, program evaluators may not have obtained participant consent for their data. Additionally, some evaluators may not seek institutional review board guidance, which can limit how data is presented or shared. We strongly encourage evaluators to seek guidance or institutional review before beginning your evaluation. For more in-depth discussions on ethics and standards of practice, we suggest Rossi et al. (2004) and Yarbrough et al. (2010) [1, 2]. Additionally, Thomas et al. (2015) provide a detailed discussion related to ethics of PE in the health professions [20].

23.5.4 Evaluation or Assessment?

In addition to differences between PE and research, there are also differences between evaluation and assessment. It is not uncommon for program managers and other stakeholders to confuse these two approaches because the terms are often used interchangeably (e.g., student evaluation versus program evaluation or program assessment compared to PE). Internationally, the term "evaluation" is usually applied at the level of a program, while "assessment" is applied to an individual [1, 2]. Importantly, although program managers or evaluators may utilize student assessment data, the goal of evaluation is to examine the program's impact on its stakeholders, which can include students. Student assessment is primarily focused on a single student.

23.6 Conclusions

The purpose of this chapter was to provide a practical overview of systematic processes of PE and advice about how to get started. PE within and outside surgical education is a widely accepted approach used to examine the efficacy of a program, determine its impact on the designated stakeholders (e.g., students, residents, patients), and to ascertain if there are any unintended consequences. Additionally, PE within surgical education can provide program managers, program directors, and other key stakeholders with important information about how students, residents, and fellows are performing, developing, and even changing their clinical practice. Important stages of a PE include identifying and working with stakeholders, who can play an integral role in focusing the evaluation's goals and questions. Including stakeholders early and staying in touch with them is a key factor in making sure the PE adds value. Furthermore, to make the PE process easier, selection of an evaluation model can provide structure, guidance, and support while helping to ensure that you do not miss important steps along the way. Lastly, although the processes of PE can be complex, there are several resources available to help guide you: many of which we have included in this chapter.

References

1. Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach*. Thousand Oaks: Sage.
2. Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users.* Los Angeles: Sage.
3. Stufflebeam, D. L. (1971). *Educational evaluation and decision making*. Ithaca: Peacock.
4. Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
5. Kring, D. (2008). Research and quality improvement: Different processes, different evidence. *MEDSURG Nursing, 17*(3), 162–169.
6. Rozalis, M. L. (2003). Evaluation and research: Differences and similarities. *The Canadian Journal of Program Evaluation, 18*(2), 1–31.
7. Lovato, C., & Wall, D. (2014). Programme evaluation: Improving practice, influencing policy and decision-making. In Swanwick, T. (Ed.), *Understanding medical education: Theory and practice* (2nd ed.). Hoboken: Wiley.
8. Gomez, P. P., Willis, R. E., & Jaramillo, L. A. (2014). Evaluation of a dedicated, surgery-oriented visiting international medical student program. *Journal of Surgical Education, 71*(3), 325–328.
9. Yu, A. Y., Wang, Q. M., Li, J., Huang, F., & Golnik, K. (2016). A cataract surgery training program: 2-year outcome after launching. *Journal of Surgical Education, 73*, 761–767.
10. Dick, W., & Carey, L. (2011). *The systematic design of instruction*.
11. Greene, J. C. (2005). Mixed methods. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks: Sage.
12. McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your programs performance story. *Evaluation and Program Planning, 22*(1), 65–72.
13. Shakman, K., & Rodriguez, S.M. (2015, May). *Logic models for program design, implementation, and evaluation: Workshop toolkit*. REL 2015-057. Regional Educational Laboratory Northeast & Islands.
14. Lawton, B., Brandon, P.R., Cicchinelli, L., Kekahio, W. (2014, February) *Logic models: A tool for designing and monitoring program evaluations*. REL 2014-007. Regional Educational Laboratory Pacific.
15. Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler Publishers.
16. Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks: Sage Publications.
17. Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications* (1st ed.). San Francisco: Jossey-Bass.
18. Torbeck, L., Canal, D. F., & Choi, J. (2014). Is our residency program successful? Structuring an outcomes assessment system as a component of program evaluation. *Journal of Surgical Education., 71*(1), 73–78.
19. Fitzpatrick, J., Sanders, J., & Worthen, B. (2011). *Reporting evaluation results: Maximizing use and understanding. program evaluation: Alternative approaches and practical guidelines* (4th ed., pp. 453–489). Upper Saddle River: Pearson Education.
20. Thomas, P. A., Kern, D. E., Hughes, M. T., & Chen, B. Y. (2015). *Curriculum development for medical education: A six-step approach*. JHU Press.