

CONCEPT PAPER

Program evaluation: An educator's portal into academic scholarship

Shera Hosseini PhD¹ | Yusuf Yilmaz PhD^{2,3}  | Kaushal Shah MD⁴ |
Michael Gottlieb MD⁵  | Christine R. Stehman MD⁶ | Andrew K. Hall MD MEd^{7,8}  |
Teresa M. Chan MD, MHPE^{9,10,11,12} 

¹Faculty of Health Sciences, McMaster Institute for Research on Aging, McMaster Education Research, Innovation, and Theory, McMaster University, Hamilton, Ontario, Canada

²McMaster Education Research, Innovation and Theory (MERIT) Program & Office of Continuing Professional Development, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

³Department of Medical Education, Faculty of Medicine, Ege University, Izmir, Turkey

⁴Department of Emergency Medicine, Weill Cornell Medical School, New York, New York, USA

⁵Department of Emergency Medicine, Rush University Medical Center, Chicago, Illinois, USA

⁶Department of Emergency Medicine, University of Illinois College of Medicine - Peoria/OSF Healthcare, Peoria, Illinois, USA

⁷Department of Emergency Medicine, University of Ottawa, Ottawa, Ontario, Canada

⁸Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada

⁹Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

¹⁰Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, Ontario, Canada

¹¹McMaster Program for Education Research, Innovation, and Theory (MERIT), McMaster University, Hamilton, Ontario, Canada

¹²Department of Health Research Methodology, Impact, and Evidence, McMaster University, Hamilton, Ontario, Canada

Correspondence

Teresa M. Chan, MD, MHPE, 237 Barton St. E., McMaster Clinics, Room 255, Hamilton, ON, L8P 1H8, Canada.
Email: teresa.chan@medportal.ca

Abstract

Program evaluation is an “essential responsibility” but is often not seen as a scholarly pursuit. While Boyer expanded what qualifies as educational scholarship, many still need to engage in processes that are rigorous and of a requisite academic standard to be labelled as scholarly. Many medical educators may feel that scholarly program evaluation is a daunting task due to the competing interests of curricular change, remediation, and clinical care. This paper explores how educators can take their questions around outcomes and efficacy of our programs and efficiently engage in education scholarship. The authors outline how educators can examine whether training programs have a desired impact and outcomes, and then how they might leverage this process into education scholarship.

INTRODUCTION

Program evaluation has been referred to as an “essential responsibility” for those tasked with the oversight of medical training programs,¹ but it is striking how little of this program evaluation work

is labelled as scholarly, and how rarely this work translates into academic scholarship. While what qualifies as educational scholarship has been expanded well beyond traditional peer-reviewed publications to include the scholarship of teaching, discovery, integration, and application,² there is still a need to engage in

processes that are rigorous and of a requisite academic standard to be labeled as scholarly.³ However, being asked to both create educational deliverables and innovate within this context is often already above and beyond the duties of overworked and under-supported medical educators. Many medical educators may feel that scholarly program evaluation is a step too far—with so many competing interests, it can be difficult to find the “bandwidth” to accomplish these scholarly tasks.⁴ Don't we all wonder about the outcomes and efficacy of our programs? Were our programs received as they were intended? And finally, is my training program having the desired impact and outcomes? And if so, wouldn't it be nice to generate a multiple win around your project?⁵

It is not just a lack of time that can prevent medical educators from engaging in scholarly evaluation efforts. Some educators may also feel inadequately trained in program evaluation and unclear what approaches and strategies to employ when engaging in program evaluation. Further, if evaluation is completed well, there is often an opportunity to translate this work into scholarly outputs.

The goal of this paper was to accomplish three goals: (1) to introduce educators to the concept of program evaluation, (2) to help them to understand frameworks that will guide them in correctly and rigorously performing program evaluations, and (3) to discuss ways in which program evaluation can translate to scholarly output.

WHAT IS PROGRAM EVALUATION?

In medical education, a “program” can refer to a large spectrum of activities, and experiences—they can range from a new workplace-based assessment program^{6,7} to a boot camp series⁸ to a longitudinal faculty development course.^{9,10} It is an ever-evolving field with new technologies, shifting paradigms, and often unclear scholarly formats. The delivery of medical education requires the implementation of programs. Whether it is a well-established program (e.g., intern orientation or airway management training) or a novel approach

to assessment (e.g., simulation-based critical care competency or entrustable professional activity), these programs need to be evaluated to determine if they are worthwhile with respect to effectiveness or value. A formal definition for program evaluation has been put forth by Mohanna and Cottrell as “a systematic approach to the collection, analysis, and interpretation of information about any aspect of the conceptualization, design, implementation, and utility of educational programmes”.¹¹ Simply stated, program evaluation is the process of identifying the value of an educational offering, but at times it can also be a way of determining issues or problems in need of systematic improvement.

Methods similar to those employed by experimentalists or epidemiologists may be used for measurement and analysis when conducting program evaluation, but this process is distinct from conventional research studies. Experimental research typically focuses on the generation of new knowledge that adds to the world more transferable or generalizable to other contexts, whereas program evaluation seeks to understand the efficacy of a specific, discrete project (e.g., a curricular change in a program or a new course design). Quantitative experiments may involve hypothesis testing with a control group and an experimental group, while qualitative studies may seek to understand or describe an experienced phenomenon. Despite being distinct from research, program evaluation is a rigorous process that might use a variety of quantitative and/or qualitative data to determine the value of the outcomes of a program, though technically a research protocol is not required.

WHY AND WHEN TO USE PROGRAM EVALUATION

While the specific purposes of program evaluation are extensive, at its core, program evaluation is about values, judgements, decision making, and change.^{1,12,13} Program evaluation is another way, outside of the program itself, that you can create a value proposition to your community via your program.¹⁴ Educators use program

TABLE 1 Purposes for program evaluation

Accountability	Knowledge	Development
<ul style="list-style-type: none"> • Was the program effective?• Was the program successful (did it meet its objectives and expectations, at the desired cost and with the desired efficiency)? • To satisfy external requirements (from funding source or accrediting body) • Are resources being appropriately allocated (including time, staff, and money)? • Can the program achieve its goals more economically? 	<ul style="list-style-type: none"> • To understand how the program is working (what were the outcomes?)• Does the result of the program derive directly from the program itself (outcome) or from the program's interactions with outside forces (impact)? • Does this program change the way that this topic/subject is taught? • What barriers to this type of program exist and how have they been overcome here? 	<ul style="list-style-type: none"> • What needs to change about the program?• How can the program be improved upon? • Feedback to the program faculty (can be used for promotion and career development) • Feedback to the program's administrative support staff (maintain or increase the support) • Does a program need to exist (needs assessment)? • Were the teaching methods used appropriate for the desired outcomes? • Program monitoring - is the program continuing to work as well as it had?

*Table derived from Woodward 2002; Thomas et al 2016; Goldie 2006; Frye & Hemmer 2012; Battista et al 2019; Durning, Hemmer & Pangaro 2007; Chelimsky & Shadish 1997.^{1,8,10-13}

evaluation to determine the value and worth of the program they designed and then explain that worth to others. There are multiple program evaluation frameworks, and which framework you select is determined by the stakeholders and focus of the evaluation.^{13,15}

The ultimate *why* of your program evaluation will be how you define success of the program in the eyes of the stakeholders and the focus of the evaluation.¹⁶ This marker of success should fall into at least one broad category of program evaluation—accountability, knowledge, or development—though these categories are often intertwined.^{1,12,17} More specific purposes for evaluation within these three categories are found in [Table 1](#).

Although it can resemble research (e.g., experimental or qualitative medical education research), it is differentiated from research by the fundamental underlying impetus for the study—research work seeks to understand the world better through its conduct (to

create generalizable or transferrable “truths” to better understand how things work), whereas program evaluation seeks to understand how and if a specific program works.

If done correctly, program evaluation is a systematic method of answering questions about the program you have designed, providing insights for others to replicate or avoid in their own programs.¹⁸ Once the work has been done, “dissemination to the community at large constitutes a critical element of scholarship.”¹³ Dissemination of this work could be publishing the program evaluation as an original research report, as an innovation report, or in an online curricular repository (e.g., MedEdPORTAL, JETem) to help advance knowledge for others ([Table 2](#)).

Overall, once the rationale is determined, program evaluation can be divided up into two groups that help direct the *when*—formative (i.e., used to improve the performance of the program,

TABLE 2 Comparing and contrasting of various types of program evaluation scholarship

	Curriculum package (eg. JETem.org or MedEdPortal)	Innovation report	Original article (Formal program evaluation study)
Prototypical Study Question	Is our program worth repeating in other contexts by other teachers?	Usually one (or a combination of) the following questions: <ul style="list-style-type: none"> - What is the theory and science that fuels, powers, and underpins this innovation? - Is our innovation useful? - Does implementing this innovation achieve some desired outcome? 	Usually seeking to ask a study question that clarifies, explains, or justifies a program. Study questions can come in a wide variety, but center upon the specific aspects of a program <ul style="list-style-type: none"> - What is the experience of participants in a specific educational program? - Does the cost of a particular program justify its existence? - What is the dropout rate for a program? - Did the program have a substantial impact that is potentially generalizable?
Description of the origins and development of the Innovation	Emphasized slightly more to explain the gap that the curricular package fills	Emphasized heavily on the actual building of the innovation Analogous to a technical report (engineering) or early materials development work (chemistry or other sciences) Theory and conceptual frameworks are often highlighted	Deemphasized May even cite the prior innovation report like a full study cites a protocol
Description of the actual Innovation	The featured element within this type of scholarship. Really details the innovations	This is certainly highlighted in some depth, but not to the level of a curricular package. May wish to append curricular materials within the appendix, but certainly NOT the center point for this type of paper	Deemphasized but usually is described with enough rigor in the materials section of the methods for a new reader (who has not yet read prior work on the topic) can understand the nature and high-level specifics of the innovation—at least so as to understand why the outcomes were of interest
Outcomes reporting	Increasingly desired but also usually provides insights to other teachers seeking to implement this curriculum as to why this is important. Usually some level of outcomes reporting (e.g., Kirkpatrick level 1, acceptability) is required	Some level of reporting for outcomes	Depending on the framing of the article the outcomes may be different from a simple reporting of effectiveness. Often original works that explore innovations will delve

program monitoring, happens at various times) and summative (i.e., used for overall judgements about the program and its developers, usually at the end of the program).^{19,20} No matter what the why, all programs should have program evaluations built into them. In fact, Woodward argues that program evaluation should be done within every part of the educational intervention process. For example, a needs assessment is the program evaluation determining the need for the program.¹⁹ Ideally, the program evaluation should be developed alongside the program itself ensuring that one does a credible evaluation answering all required questions.¹⁸ Early program evaluation development prevents later problems and allows data to be collected, as suggested by Durning et al.,¹⁶ during three phases: (1) before (establish a baseline and helps show how much of the outcomes are due to the program itself), (2) during (process measurements; allows developers to notice and fix problems early), and (3) after the program (outcome measurements). The why and when of program evaluation feed directly into the approach you take in doing the program evaluation (i.e., how you actually do this).

HOW TO USE PROGRAM EVALUATION METHODOLOGIES

As stated above, development of the program evaluation should happen alongside development of the program itself, meaning *prior* to launching the program (or the most recent class of participants). This involves identifying the specific goals of the evaluation by considering the potential stakeholders and end-users of the resultant evaluation. With this information, educators can better align the breadth and focus of the evaluation with their specific needs (Box 1).

Once you have identified the target audience, next determine the underlying theory for change. The three most common theories for this are reductionism, system theory, and complexity theory. Reductionism relies upon an assumption that there is a specific order with a direct cause and effect for each action.²¹ This approach, reflected in models such as the Logic model,^{22,23} suggests that there is a clear linearity and predictable impact from each intervention.¹ System theory builds upon this with its roots in the general system theory applied to biology.²⁴ In this model, it

is proposed that the whole of a system is greater than the sum of its individual parts.²⁴ Therefore, education programs expand beyond merely isolated parts, instead comprising the integration of the specific program components with each other and with the broader educational environment. Complexity theory expands further to adapt to the ever-changing, more complex state of programs in real life.^{1,25} There are multiple complex factors that can influence education programs, including the participants, influence of stakeholders and regulators, professional practice patterns, the surrounding environment, and expanding knowledge within the specific field, as well as with regard to the education concepts being taught.¹ Understanding the underlying theories can help inform the conceptual frameworks selected for evaluation, but we will dive into this more in the next section.

CONCEPTUAL FRAMEWORKS

There are many frameworks that can guide your program evaluation process. A full description of each of these is beyond the scope of this paper; however, our authorship team has detailed six program evaluation frameworks that have been featured in medical education (and specifically AEM Education and Training) including: CIPP, Kirkpatrick Model, Logic Model, Realist Evaluation, RE-AIM, and SQUIRE-EDU. Table 3 provides a description of some of the more commonly used frameworks and sources of further information on each of them.³¹⁻⁴⁷

When creating the program evaluation, you may utilize frameworks to guide the data collection. The selection process for your conceptual framework will require consideration of the end-users and which data will be most valuable to them. You should perform a thorough literature search to identify similarities and differences with prior programs. Questions should seek to assess the benefits and consequences of the new intervention or innovation. During the literature search, seek out existing tools used by similar programs to inform your evaluation tool design. Identify how this aligns with your current program evaluation needs and modify the tool where necessary. It is important to also collect validity evidence for your specific tool.²⁶ Even if a tool is “validated” in another setting, new validity should be sought for the current application within the context of the new program.²⁶ Since evaluation is often centered on a particular program, the evaluation plan may contain outcomes that are idiosyncratic rather than generalizable; however, best practices of questionnaire design should still be followed as much as possible (e.g., basing the tool on prior evaluation of a previous study, pilot testing a survey tool prior to launch to ensure readability and clarity).

Finally, consider the outcomes with a broader lens. While often considered with regard to learner-oriented outcomes (e.g., Kirkpatrick model), it is also important to consider the costs (e.g., time, expenses, faculty) and broader societal implications as described further below. Those reading the findings will want to weigh the cost and benefits of the program.

BOX 1 Components of a program evaluation

1. Develop an evaluation question based on specific goals of various stakeholders
2. Identify your theory of change
3. Perform a literature search
4. Identify your (validated) collection instrument
5. Consider your outcomes with a broad lens

TABLE 3 Conceptual frameworks for program evaluation

Framework name	Origins & explanation	Example scenario	Citation for a "How to" guide (first author, year)	Citation of an exemplar paper (first author, year)
CIPP (Context, Input, Process, Products)	<p>CIPP is a comprehensive framework for guiding the evaluations of programs and systems and includes the following components: context, input, process, and impact</p> <p>Context is based on needs assessment, available resources, problems, any background information, and the overall program environment. It includes the planning stage and mainly focuses on the desired goals and objectives for a program</p> <p>Inputs refer to the required strategies, tools, or resources that must be included in the program to meet the needs identified during the Context stage. Inputs include elements including budget, research, plans, stakeholders, or subject matter experts</p> <p>Process is the stage of program development and execution.</p> <p>This stage is where the inputs all come together and it is often revisited to ensure that the program development was well-designed and that the program implementation is meeting expectations</p> <p>Products include the review phase and are the outputs and outcomes related to program performance and objectives</p> <p>The main question in this stage is whether the intended goals have been met. Further, the program sustainability in terms of context, inputs and processes as well as any potential necessary changes to the program are assessed</p>	<p>The program director recently launched a new diversity, equity, and inclusion curriculum. She wants to better understand and evaluate the effectiveness of the curricula. She selects the CIPP framework to better understand the context, inputs, process, and products, so as to more fully consider all of the inputs and outputs from the new curriculum</p>	<p>Stufflebeam (2003)²¹ Lee (2019)²²</p>	<p>Steinert (2005)²³ Rooholamini (2017)²⁴</p>
Kirkpatrick Model	<p>Kirkpatrick's original four-level (reaction, learning, behavior, results) model is widely employed in the evaluation of health professional education programs</p> <ol style="list-style-type: none"> 1. Reaction: Assesses a person's reactions to a course related element such as teachers, materials, activities, and design While high satisfaction does not necessarily guarantee the next level (learning), low satisfaction levels are likely to reduce the probability of learning 2. Learning: refers to changes in knowledge, skills, and attitudes 3. Behaviors: are changes in practice 4. Results: changes at the organizational level <p>Other more recent modifications of this model include:</p> <ul style="list-style-type: none"> • The New World Kirkpatrick model; • Phillips (Includes Return on Investment (RoI)); • Kaufman (adds Impact on Society); • Barr et al. and Hammick et al. (include impact of interprofessional education) 	<p>The vice chair of faculty development has created a new asynchronous, just-in-time training module. She wants to understand the perception and effect of this new module. As part of the evaluation, she sought out users' reaction, learning, behaviors, and impact on the system. She selected the Kirkpatrick framework to ensure she had both perception and higher-level outcomes</p>	<p>Kirkpatrick (2006)²⁵ Kirkpatrick and Kirkpatrick (2016)²⁶ Barr (2005)²⁷ Hammick (2007)²⁸ The New World Kirkpatrick Model²⁹ Phillips (2003)³⁰ Kaufman and Keller (1994)³¹</p>	<p>Gottlieb (2021)³² Lam and Stickrath (2020)³³</p>
Logic Model	<p>This model is commonly used for designing and evaluating projects and consists of a matrix that outlines a project's goals, activities, assumptions, and expected results. It provides a structure to help clarify the components of a project, the activities, resources, as well as its anticipated challenges</p>	<p>The medical student clerkship director has added a new airway curriculum for the medical students on rotation. He realizes that it is important to understand both the cost and benefits of the program. Therefore, he uses the Logic model to incorporate both the inputs and outputs into the program evaluation</p>	<p>Newcomer (2015)³⁴ Van (2016)¹⁷</p>	<p>Love (2016)³⁵</p>

TABLE 3 (Continued)

Framework name	Origins & explanation	Example scenario	Citation for a "How to" guide (first author, year)	Citation of an exemplar paper (first author, year)
Realist Evaluation	Realist evaluation is suited for the evaluation of complex educational interventions such as simulation-based education. It seeks to answer what works, for whom, in what circumstances, in what respects, to what extent, and why It uses a mixed methods approach to collection of data to test the context-mechanism-outcome configurations of the education intervention Through investigating the context, mechanisms, and outcomes of education programs, realist evaluation can allow educators to better understand why and when an evaluation does work and in which contexts	The simulation director has created a new in-situ simulation program which includes interprofessional learners from multiple professions. He wants to identify what works best for different learners in different circumstances and why. Therefore, he selects a realist evaluation for his framework	Graham and McAleer (2018) ³⁶ Wong (2012) ³⁷	Ogrinc (2014) ³⁸ Ellaway (2018) ³⁹
RE-AIM	The RE-AIM framework is mainly designed to evaluate the impact of community-based public health programs and interventions. These interventions are often complex as they mainly rely on multiple stakeholders and in complex settings. To understand the impact of a program, impact on participants, organization providing a program, and the broader community needs to be captured This framework consists of five evaluation dimensions: Reach, Effectiveness, Adoption, Implementation, and Maintenance. It has been implemented across various settings and contexts such as community, policy, and public health initiatives	The vice chair of operations created a new program to train their physicians, advanced practice providers, and nurses on using telemedicine for patient care. Given the complexity of the intervention and reliance upon multiple stakeholders, she uses the RE-AIM framework	Glasgow (1999) ⁴⁰ Shaw (2019) ⁴¹	Nagji (2020) ⁴² Rose (2021) ⁴³ Yilmaz (2021) ⁴⁴
SQUIRE-EDU	The SQUIRE (Standards for Quality Improvement Reporting Excellence) framework offers guidelines for reporting new knowledge about ways to improve healthcare. SQUIRE has an adaptation that is specifically about Educational Improvement (SQUIRE-EDU) These guidelines are proposed for reports on system-level work to improve the quality, safety, and value of healthcare systems SQUIRE offers a variety of ways to improve healthcare and encourages the researchers to consider all SQUIRE items, but the inclusion of every SQUIRE element may not be necessary SQUIRE has provided guidance on healthcare improvement, and contributed to the understanding of factors that impact the success, and failure, of healthcare improvement efforts	The medical director is leading a quality improvement initiative to reduce overprescribing of antibiotics. He has developed a multi-step program which includes a specific training session. He uses the SQUIRE-EDU framework to align with the focus on quality improvement	Goodman (2016) ⁴⁵ Ogrinc, (2019) ⁴⁶	Taylor (2019) ⁴⁷

MARKERS OF HIGH-QUALITY PROGRAM EVALUATION

Program evaluation and research studies have very common features, depending on the objectives of a study, these two methods may become very similar. While research studies aim to produce new knowledge, program evaluation studies focus on the program quality and value.²⁷ When unsure, ethics boards guidelines are helpful for ensuring that the study that you are about to conduct is a program evaluation study. In the United States, many program evaluations will require institutional review board approval but are usually granted exemption status since program evaluations will fall well within normal educational practices. Ethics boards in Canada deem program evaluations exempt from the ethical review as per Tri-Council Policy Statement 2 [2018] Article 2.5.²⁸ Therefore, initially, a program evaluation study should be checked with the ethics

BOX 2 Pearls for those interested in conducting program evaluation work

Based on prior literature on innovation reports and program evaluations, we have identified some common problems encountered when authors claim to have conducted these formats of studies:

Pearl 1: Plan the program evaluation from the onset. Ideally, program evaluation should be established prior to the program launch (or at least prior to the most recent cohort). Performing program evaluation once the program is ongoing will limit the available information and increase the risk of recall bias.

Pearl 2: Consider all of the inputs and outputs. The evaluators will need to think beyond just the learner outcomes and consider the broader outcomes, impacts, and the resources and requirements to run the program.

Pearl 3: Attempt to identify unintended outcomes. Intended outcomes are often tracked but a systematic inquiry into identifying unintended outcomes is often overlooked.

Pearl 4: Involve a statistician or a data scientist early. Some program evaluation approaches require complex statistical analysis and even further data exploration to understand complex data to be collected through the program implementation. A statistician or a data scientist can provide different approaches on how to analyze data and understand the relationship on program focus and outcomes.

Pearl 5: Chart the overall program evaluation process. Program evaluation could be very complex from planning to evaluation. Each step of the program evaluation should be represented with a figure in the study. This charting process will give readers a clear idea about the program evaluation steps and how the framework was implemented at each step.

board and receive an ethical exemption to make sure that the study purpose, objectives, data collection, and analysis aligns with it.

There are three common approaches to program evaluation studies: decision-oriented, outcomes-oriented, and expertise-oriented.²⁹ In the previous section, various program evaluation frameworks and models were described that can yield to the overall

BOX 3 Key resources for further reading

The following are key papers on the program evaluation methodology recommended for those interested in learning more.

1. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. *Med Teach*. 2012;34(5):e288-e299.

This is a review of several common program evaluation models and the benefits and limitations of each. The paper also provides examples of how to apply these in practice.

2. Cook DA. 2010. Twelve tips for evaluating educational programs. *Med Teach*. 32:296–301.

A concise article that breaks down program evaluation into twelve “tips” to guide the development and implementation. Not meant to be used alone, but again a solid introduction to the process with an included blank table for readers to start brainstorming their own program evaluations.

3. Goldie J. AMEE Education Guide no. 29: Evaluating Educational Programs. *Med Teach*. 2006; 28(3): 210–224.

An introductory how-to guide for program evaluation of educational programs in general including the history and the process. A solid starting point for someone who is unfamiliar with the process and a solid introduction to allow better integration of the information provided in the AMEE no. 67 (included below) which walks the reader through theories to use as frameworks for their program evaluations.

4. Durning SJ, Hemmer P, Pangaro LN. The Structure of Program Evaluation: An Approach for Evaluating a Course, Clerkship, or Components of a Residency or Fellowship Training Program. *Teach Learn Med*, 19:3, 308–318, [10.1080/10401330701366796](https://doi.org/10.1080/10401330701366796)

While the other articles included here involve program evaluation in general, this article focuses on applying program evaluation to graduate medical education. While it is just one particular framework out of many that are available, it provides insight into how to apply program evaluation to programs that don't necessarily fit the usual educational program mold. For medical educators beginning their program evaluation journey, having this example will allow them to see how other frameworks might be used for their programs.

approaches. These frameworks are of vital value to the overall program evaluation process.¹ Without using a framework, program evaluation may lose its focus and the flow of the study may become redundant and less helpful. As each framework focuses on different parts of a study, it is important for researchers to take into account the study's objectives and focus. The face validity of a framework should be agreed by the investigators, meaning the outcomes of the study could be achieved through the selected framework.¹³ A study could focus on many objectives such as trainees' learning, satisfaction, and the intervention's success in reaching various audiences.¹

Innovation reports are an integral part of program evaluation studies as they evaluate novel approaches to teaching and learning. Hall and colleagues reviewed the literature on the quality markers of innovation reports and came up with 34 items resulting in seven themes from analysis of the problem to dissemination of results to ensure that the innovation reports adequately provide insights and reproducibly.³⁰ Therefore, ensuring that a program evaluation study has rigor and reproducibility is very important for any type of program evaluation study. **Box 2** provides various pearls to help researchers who will tackle to program evaluation studies. **Box 3** contains an annotated bibliography that summarizes key resources for further reading.

CONCLUSION

Program evaluations can be seen as a gateway towards other forms of scholarship for those who are most at home developing programs and curricula. However, it should be acknowledged as its own form of scholarship that is unique and separate from curriculum development or research.

CONFLICTS OF INTEREST

Dr. Shera Hosseini has received funding for her postdoctoral fellowship from the McMaster Institute for Research in Aging (MIRA). Dr. Yilmaz is the recipient of a 2019 TUBITAK Postdoctoral Fellowship grant. Dr. Shah—none; and no grants. Dr. Gottlieb holds grants for unrelated work with the Centers for Disease Control and Prevention, Council of Residency Directors in Emergency Medicine, Society for Academic Emergency Medicine, and eCampus Ontario. Dr. Stehman—none, Dr. Hall—holds grants for unrelated work from the Royal College of Physicians and Surgeons of Canada, Queen's University Center for Teaching and Learning, and the Physician Services Incorporated Foundation. Dr. Chan holds grants for unrelated work from McMaster University, the PSI foundation, Society for Academic Emergency Medicine, eCampus Ontario, the University of Saskatchewan, and Royal College of Physicians and Surgeons of Canada.

ORCID

Yusuf Yilmaz  <https://orcid.org/0000-0003-4378-4418>

Michael Gottlieb  <https://orcid.org/0000-0003-3276-8375>

Andrew K. Hall  <https://orcid.org/0000-0003-1227-5397>

Teresa M. Chan  <https://orcid.org/0000-0001-6104-462X>

REFERENCES

1. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE Guide No. 67. *Med Teach* 2012;34(5):e288-e299. doi:[10.3109/0142159X.2012.668637](https://doi.org/10.3109/0142159X.2012.668637)
2. Boyer EL. The Scholarship of Engagement. *Bull Am Acad Arts Sci*. 1996;49(7):18-33. doi:[10.2307/3824459](https://doi.org/10.2307/3824459)
3. Van Melle E, Lockyer J, Curran V, Lief S, St Onge C, Goldszmidt M. Toward a common understanding: supporting and promoting education scholarship for medical school faculty. *Med Educ*. 2014;48(12):1190-1200. doi:[10.1111/medu.12543](https://doi.org/10.1111/medu.12543)
4. Zibrowski EM, Weston WW, Goldszmidt MA. "I don't have time": issues of fragmentation, prioritisation and motivation for education scholarship among medical faculty. *Med Educ*. 2008;42(9):872-878. doi:[10.1111/j.1365-2923.2008.03145.x](https://doi.org/10.1111/j.1365-2923.2008.03145.x)
5. Gottlieb M, Chan TM, Sherbino J, Yarris L. Multiple wins: embracing technology to increase efficiency and maximize efforts. *AEM Educ Train*. 2017;1(3):185-190. doi:[10.1002/aet2.10029](https://doi.org/10.1002/aet2.10029)
6. Chan T, Sherbino J. The McMaster modular assessment program (McMAP). *Acad Med*. 2015;90(7):900-905. doi:[10.1097/ACM.0000000000000707](https://doi.org/10.1097/ACM.0000000000000707)
7. Li S, Sherbino J, Chan TM. McMaster modular assessment program (McMAP) through the years: residents' experience with an evolving feedback culture over a 3-year period. *AEM Educ Train*. 2017;1(1):5-14. doi:[10.1002/AET2.10009](https://doi.org/10.1002/AET2.10009)
8. McMurray L, Hall AK, Rich J, Merchant S, Chaplin T. The nightmares course: a longitudinal, multidisciplinary, simulation-based curriculum to train and assess resident competence in resuscitation. *J Grad Med Education*. 2017;9(4):503-508. doi:[10.4300/JGME-D-16-00462.1](https://doi.org/10.4300/JGME-D-16-00462.1)
9. Chan TM, Gottlieb M, Sherbino J, et al. The ALiEM faculty incubator: a novel online approach to faculty development in education scholarship. *Acad Med*. 2018;93(10):1497-1502. doi:[10.1097/ACM.0000000000002309](https://doi.org/10.1097/ACM.0000000000002309)
10. Gottlieb M, Yarris LM, Krzyzaniak SM, et al. Faculty development using a virtual community of practice: three year outcomes of the academic life in emergency medicine faculty incubator program. *AEM Educ Train*. 2021;5(3):e10626. doi:[10.1002/aet2.10626](https://doi.org/10.1002/aet2.10626)
11. Mohanna K, Cottrell E. *Teaching made easy: a manual for health professionals*. Abingdon, UK: Radcliffe Publishing; 2010.
12. Goldie DJ. AMEE education guide no. 29: evaluating educational programmes. *Med Teach*. 2006;28(3):210-224. doi:[10.1080/01421590500271282](https://doi.org/10.1080/01421590500271282)
13. Cook DA. Twelve tips for evaluating educational programs. *Med Teach*. 2010;32(4):296-301. doi:[10.3109/01421590903480121](https://doi.org/10.3109/01421590903480121)
14. Battista A, Yoon M, Matthew Ritter E, Nestel D. Demystifying program evaluation for surgical education. In: Nestel D, Dalrymple K, Paige JT, Aggarwal R, eds. *Advancing surgical education*. Springer Singapore; 2019:255-267. doi:[10.1007/978-981-13-3128-2_23](https://doi.org/10.1007/978-981-13-3128-2_23)
15. Simpson D, Riddle JM, Hamel DL Jr, Balmer DF. Blueprinting program evaluation evidence through the lens of key stakeholders. *J Grad Med Educ*. 2020;12(5):629-630. doi:[10.4300/JGME-D-20-01015.1](https://doi.org/10.4300/JGME-D-20-01015.1)
16. Durning SJ, Hemmer P, Pangaro LN. The structure of program evaluation: an approach for evaluating a course, clerkship, or components of a residency or fellowship training program. *Teach Learn Med*. 2007;19(3):308-318. doi:[10.1080/10401330701366796](https://doi.org/10.1080/10401330701366796)
17. Chelmsky E, Shadish WR. *Evaluation for the 21st century: a handbook*. Thousand Oaks, CA: Sage; 1997. doi:[10.4135/9781483348896](https://doi.org/10.4135/9781483348896)
18. Lovato C, Peterson L. *Programme Evaluation*. In: *Understanding Medical Education*. John Wiley & Sons, Ltd; 2018:443-455. doi:[10.1002/9781119373780.ch30](https://doi.org/10.1002/9781119373780.ch30)
19. Woodward CA. Program evaluation. In: *International handbook of research in medical education*. Cham, Switzerland: Springer; 2002:127-155. doi:[10.1007/978-94-010-0462-6](https://doi.org/10.1007/978-94-010-0462-6)
20. Thomas PA, Kern DE, Hughes MT, Chen BY. *Curriculum development for medical education: a six-step approach*. Baltimore, MD: Johns Hopkins University Press; 2016.

21. Stufflebeam DL, Coryn CL. *Evaluation theory, models, and applications*, vol. 50. Hoboken, NJ: John Wiley & Sons; 2014.
22. Frechtling JA. *Logic modeling methods in program evaluation*, vol. 5. John Wiley & Sons; 2007.
23. Van Melle E. Using a logic model to assist in the planning, implementation, and evaluation of educational programs. *Acad Med*. 2016;91(10):1464. doi:[10.1097/ACM.0000000000001282](https://doi.org/10.1097/ACM.0000000000001282)
24. Von Bertalanffy L. The history and status of general systems theory. *Acad Manage J*. 1972;15(4):407-426. doi:[10.5465/255139](https://doi.org/10.5465/255139)
25. Mennin S. Complexity and health professions education. *J Eval Clin Pract*. 2010;16(4):835-837. doi:[10.1111/j.1365-2753.2010.01502.x](https://doi.org/10.1111/j.1365-2753.2010.01502.x)
26. Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463-474. doi:[10.3109/0142159X.2014.889814](https://doi.org/10.3109/0142159X.2014.889814)
27. Adams J, Neville S. Program evaluation for health professionals: what it is, what it isn't and how to do it. *Int J Qual Methods*. 2020;19. doi:[10.1177/1609406920964345](https://doi.org/10.1177/1609406920964345)
28. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. Vol 45. December 2.; 2018.
29. Blanchard RD, Torbeck L, Blondeau W. AM last page: a snapshot of three common program evaluation approaches for medical education. *Acad Med*. 2013;88(1):146. doi:[10.1097/ACM.0b013e3182759419](https://doi.org/10.1097/ACM.0b013e3182759419)
30. Hall AK, Hagel C, Chan TM, Thoma B, Murnaghan A, Bhanji F. The writer's guide to education scholarship in emergency medicine: education innovations (part 3). *CJEM*. 2018;20(3):563-570. doi:[10.1017/cem.2017.28](https://doi.org/10.1017/cem.2017.28)
31. Kaufman R, Keller JM. Levels of evaluation: beyond Kirkpatrick. *Hum Resour Dev Quart*. 1994;5(4):371-380.
32. Gottlieb M, Yarris L, Krzyzaniak SM, et al. Faculty development using a virtual community of practice: three year outcomes of The Academic Life in Emergency Medicine Faculty Incubator Program. *AEM Educ Train*. [In Press]; 2021. doi:[10.1002/aet2.10626](https://doi.org/10.1002/aet2.10626)
33. Lam R, Stickrath C. It Takes a Village: Utilizing a community-based longitudinal integrated clerkship model at a regional medical campus to provide the core emergency medicine clerkship experience. *AEM Educ Train*. 2020;5(1):5-11. doi:[10.1002/aet2.10442](https://doi.org/10.1002/aet2.10442)
34. Newcomer KE, Hatry HP, Wholey JS. *Handbook of Practical Program Evaluation*, 4th ed. John Wiley and Sons Inc Publishers; 2015.
35. Love JN, Yarris LM, Santen SA, et al. A novel specialty-specific, collaborative faculty development opportunity in education research: program evaluation at five years. *Acad Med*. 2016;91(4):548-555.
36. Graham AC, McAleer S. An overview of realist evaluation for simulation based education. *Adv Simul*. 2018;13(3).
37. Wong G, Greenhalgh T, Westhorp G, Pawson R. Realist methods in medical education research: What are they and what can they contribute? *Med Educ*. 2012;46:89-96.
38. Ogrinc G, Ercolano E, Cohen ES, et al. Educational system factors that engage resident physicians in an integrated quality improvement curriculum at a VA hospital: a realist evaluation. *Acad Med*. 2014;89(10):1380-1385.
39. Ellaway RH, Palacios Mackay M, Lee S, et al. The impact of a national competency-based medical education initiative in family medicine. *Acad Med*. 2018;93(12):1850-1857.
40. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327.
41. Shaw RB, Sweet SN, McBride CB, et al. Operationalizing the reach, effectiveness, adoption, implementation, maintenance (RE-AIM) framework to evaluate the collective impact of autonomous community programs that promote health and well-being. *BMC Public Health*. 2019;19:803.
42. Nagji A, Yilmaz Y, Zhang P, et al. Converting to connect: a rapid REAIM evaluation of the digital conversion of a clerkship curriculum in the age of Covid-19. *AEM Educ Train*. 2020;4(4):330-339.
43. Rose CC, Haas MRC, Yilmaz Y, et al. ALiEM connect: large-scale, interactive virtual residency programming in response to Covid-19 [published online ahead of print, 2021 Apr 20]. *Acad Med*. 2021. doi:[10.1097/ACM.0000000000004122](https://doi.org/10.1097/ACM.0000000000004122)
44. Yilmaz Y, Sarikaya O, Senol Y, et al. REAIMing COVID-19 online learning for medical students: a massive open online course evaluation. *BMC Med Educ*. 2021;21:303. doi:[10.1186/s12909-021-02751-3](https://doi.org/10.1186/s12909-021-02751-3)
45. Goodman D, Ogrinc G, Davies L, et al. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, vol 2.0: examples of SQUIRE elements in the healthcare improvement literature. *BMJ Qual Saf*. 2016;25:e7.
46. Ogrinc G, Armstrong GE, Dolansky MA, Singh MK, Davies L. Standards for quality improvement reporting excellence in education: publication guidelines for educational improvement. *Acad Med*. 2019;94(10):1461-1470.
47. Taylor LM, East-Telling CL, Ellerton A. Exploring preceptorship programmes: implications for future design. *J Clin Nurs*. 2019;28(7-8):1164-1173.

How to cite this article: Hosseini S, Yilmaz Y, Shah K, et al. Program evaluation: An educator's portal into academic scholarship. *AEM Educ Train*. 2022;6(Suppl. 1):S43-S51. doi:[10.1002/aet2.10745](https://doi.org/10.1002/aet2.10745)