

American Journal of Evaluation

<http://aje.sagepub.com>

The Role of Evaluation in Research Practice Integration Working Toward the “Golden Spike”

Jennifer Brown Urban and William Trochim
American Journal of Evaluation 2009; 30; 538
DOI: 10.1177/1098214009348327

The online version of this article can be found at:
<http://aje.sagepub.com/cgi/content/abstract/30/4/538>

Published by:



<http://www.sagepublications.com>

On behalf of:

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: <http://aje.sagepub.com/cgi/alerts>

Subscriptions: <http://aje.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://aje.sagepub.com/cgi/content/refs/30/4/538>

The Role of Evaluation in Research–Practice Integration

Working Toward the “Golden Spike”

American Journal of Evaluation
30(4) 538-553
© The Author(s) 2009
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1098214009348327
<http://aje.sagepub.com>



Jennifer Brown Urban¹ and William Trochim²

Abstract

Program evaluation and planning is at the heart of efforts to integrate the domains of practice and research. Traditionally, research and practice have operated in independent spheres with practitioners focused on the implementation of programs that affect individual behavior and researchers focused on the development and testing of theory. Evidence-based practice (EBP), practice-based evidence, and translational research have attempted to unite these worlds, and although significant advances have been made, there is a continued need to find mechanisms that enable a seamless connection between knowledge generation and application. We propose a method that builds on the traditions of theory-driven evaluation, logic modeling, and systems science and uses evaluation and program planning as the bridging mechanism between research and practice. Included in this approach are methods that aid in the explicit expression of implicit theories, management of evaluation resources, and linkage of program theory and evaluation measures to a research evidence base.

Keywords

STEM, education, evaluation planning, systems, systems evaluation, pathways, evidence-based practice, theory-driven evaluation

Traditionally, research and practice have operated in independent spheres. This article argues that evaluation is situated in a key position between these realms and, therefore, evaluators play a key role in linking these two areas. Social science researchers are trained in developing and testing theories of social behavior. These theories are typically derived from published literature, rational ruminations of individual experts, and the careful and critical examination of natural phenomena (Trochim, 1985). To be qualified to develop and test their theories, scientists invest years in their education and training and, perhaps due in part to such investment, regard researcher knowledge as privileged and superior. That is, such training is deemed necessary for one to be qualified to

¹ Montclair State University, Montclair, New Jersey

² Cornell University, Ithaca, New York

Corresponding Author:

Jennifer Brown Urban, Montclair State University, Department of Family and Child Studies, 1 Normal Ave., University Hall, Room 4144, Montclair, NJ 07043. Email: urbanj@mail.montclair.edu

develop and test complex theories of human behavior. Typically, the goal of social science research is to explain human behavior on a broadly generalizable or global scale. Knowledge generation is based on the careful and systematic postulation and testing of explicit theories. However, with increased demand for accountability and knowledge sharing, research funding agencies are requiring researchers to articulate the worth of their research findings to broader society and even stake research dollars on the ability to effectively disseminate research to the broader public.

Alternatively, program practitioners are typically concerned with the practice and implementation of programs that affect and change individual behavior. Practitioner emphases are generally more local than researchers and are focused on effective implementation strategies and the appropriate matching of programs with client needs. Frequently, theories of implementation are developed implicitly and are not articulated or expressed in formal models. Knowledge of “what works” is derived from reflective practice and tacit knowledge based on the experience of working with clients. Practitioners may even adapt their practice to the changing needs of clients and base decisions and judgments not on well-articulated and tested theories but rather on grassroots knowledge. However, in the eyes of many funders and policy makers, a practitioner’s tacit knowledge of a program’s effectiveness is no longer a legitimate rationale for continuing to fund a program. As such, practitioners are faced with increasing pressures to base their work on carefully designed and tested programs that have been empirically demonstrated to be effective.

Policy makers or funders are facing increasing pressures to demonstrate the effective allocation of scarce resources to affect the greatest impact. They are responsible for funding both research and programs that work toward creating a healthier, more prosperous society. Typically, they are less concerned with what is occurring at the local program sites or in particular research labs. Rather, they are more result oriented and tend to look across a broad portfolio to understand “the big picture.”

The Systems Challenge

We contend that the primary challenge in research–practice integration is a failure to frame the effort from a systems perspective (Cabrera, 2006; Trochim, Cabrera, Milstein, Gallagher, & Leischow, 2006; Williams & Imam, 2006). Systems thinking includes a philosophical approach to science and evaluation that provides ways to handle complexity, link local and global, account for dynamic changes in the system or program, recognize the natural evolution of the program, and help identify leverage points (Williams & Imam, 2006). Evaluation is uniquely situated in this system because evaluators have a connection to both the practice and research realms. Systems evaluation is an approach to conducting program evaluation that considers the complex factors that are inherent in the larger system within which the program is embedded. Systems evaluation provides both a conceptual framework for thinking about evaluation systems and a set of specific methods and tools that enhance our ability to accomplish high-quality evaluation with integration across organizational levels and structures (Cornell Office for Research on Evaluation, 2009).

Practitioners and researchers face a unique set of challenges that systems evaluation aims to address. The practitioners’ dilemma is that they operate on a local level yet they are asked to demonstrate effects on long-term policy outcomes; and, they are expected to do so with limited resources. The researchers’ dilemma is that they are operating from a more theoretical level and the knowledge that they generate is seldom disseminated and used effectively by practitioners.

This is a classic systems thinking problem, a part–whole or local–global challenge (Young, 1999)—how do we connect the many varied local experiences with the broader global outcomes of interest? Research, practice, and policy generally operate at different levels of scale. For example, practitioners are embedded in the immediate context and are typically concerned with more

proximal issues such as the experiences of their program participants and how service delivery can be improved locally. Their time frame for knowledge acquisition is typically short-term. Researchers typically work on more distal, global questions related to the long-term impact. Their time frame for knowledge acquisition is typically long-term and driven by small, cumulative steps. Policy makers and funders are typically interested in the most long-term outcomes and they need the most rapid feedback.

Each of these stakeholders have power and influence over the other and in turn have needs that only the others can meet. Policy makers have power in terms of their ability to exert influence through the allocation of money. They decide where research dollars should be invested and what programs should be funded. However, they also need researchers and practitioners to provide data that will inform their decision making. Practitioners have power because they provide access to much of the data and influence the quality of the data collected for use by both policy makers and researchers. However, they need funding to run their programs, which they can only obtain if they are able to demonstrate the effectiveness of their programs. Researchers have the luxury of time to conduct studies focused on long-term outcomes; however, they are frequently required to disseminate their findings and apply them to real-world problems. At the core of this systems challenge is the central role that program evaluation, planning, and evaluators can play in making these local-global connections toward integrating the work of research and practice.

This article begins with a discussion of previous attempts at research-practice integration. This is followed by a discussion of the role of evaluators as facilitators in the research-practice integration process. The systems evaluation partnership (SEP) is then presented as a model for achieving research-practice integration via carefully planned and executed evaluation. This includes a description of evaluation scope analysis, evidence mapping, and measurement mapping. The article concludes with a discussion of how this approach can be supplemented and enhanced with a supporting cyberinfrastructure.

Prior Attempts at Connecting the Research-Practice Divide

There is an enormous literature addressing the relationship between research and practice, which considers how the two might be more effectively integrated and includes research on a variety of methods and approaches to dissemination and implementation (Durlak & Dupre, 2008; Galavotti, Kuhlmann, Kraft, Harford, & Petraglia, 2008; Wandersman et al., 2008) and capacity building for evidence-based approaches (Rolleri, Wilson, Paluzzi, & Sedivy, 2008). This literature goes considerably beyond the origins of the evidence movement in traditional biomedical research and encompasses such varied fields as teen pregnancy (Lesesne et al., 2008), violence prevention (Guerra & Knox, 2008; Saul et al., 2008), psychotherapy (Margison et al., 2000), and gerontology (Olson, 1996).

Evidence-Based Practice (EBP)

EBP is one attempt at unifying the worlds of research and practice. Originating in medicine (Antes, 1998; Sackett, 1997) and now migrating to many other fields (Banerjee & Dickinson, 1997; Brownson, Baker, Leet, & Gillespie, 2002; Eitel & Steiner, 1999; Gibbs, 2003; Mupparapu, 2002), EBP is grounded in the idea that programs that have a strong evidence base demonstrating their effectiveness should be disseminated and implemented. EBP unites the worlds of research and practice by helping to ensure that practice is informed by research knowledge. However, EBP tends to be framed primarily from a researcher, not a practitioner perspective and prioritizes knowledge generation over practical problems, and precision and control over generalizability and diffusion (Green, 2007).

Practice-Based Evidence

A more recent response to the disproportionate balance between researcher versus practitioner generated knowledge inherent in EBP has been the discussion and promotion of what can be termed practice-based evidence (Green, 2006; McDonald & Viehbeck, 2007; Westfall, Mold, & Fagnan, 2007). According to this approach, research agendas are derived from and responsive to the needs identified in practice. Here, the impetus for knowledge acquisition consists of the questions raised by the practitioners who have daily experience interacting with clients, identifying needs, and generating “theories” of best practice.

Translational Research

Although the goal of both EBP and practice-based evidence is the integration of research and practice, these approaches have, thus far, failed to adequately unite the two because they lack a clear bidirectional process that does not privilege one over the other. Like EBP, translational research (Dougherty & Conway, 2008; Khoury et al., 2007; Sung et al., 2003; Westfall et al., 2007; Woolf, 2008)—which has been widely applied in biomedicine, mental health, and public health—attempts to bridge the gap from bench to bedside or from research to practice. There is broad consensus in the literature that translational research is bidirectional and dynamic in nature (Busse & Fleming, 1999; Cesario, Galetta, Russo, Margaritora, & Granone, 2003; Horig & Pullman, 2004; Ter Linde & Samson, 2004; Westfall et al., 2007). Although translational research has made great strides toward integrating basic research with clinical practice, the advances have primarily been anchored in the biomedical world and are focused on treating the individual in a clinical setting—exemplified by the phrase “bench to bedside.” Similar advances have yet to be made for programs in nonclinical settings (e.g., community-based or school-based programs) or for programs that emphasize behavior change and/or education.

Evaluators as Facilitators for Research–Practice Integration

Although significant advances have been made toward uniting research and practice via EBP, practice-based evidence, and translational research, there is a continued need to work at finding mechanisms that enable a more seamless connection between knowledge generation and the application of that knowledge to improve the human condition. Evaluators have the tools, and indeed a responsibility, to help achieve a greater connectivity between research and practice, particularly for programs aimed at behavioral, social, and/or educational change.

In particular, evaluators provide the link between research and practice through their role in facilitating the development of visual causal diagrams. Causal modeling has a history in both practice (i.e., theory of change) and research (i.e., causal path modeling). The field of system dynamics has always had an emphasis on causal modeling, especially models that incorporate causal feedback loops (Richardson, 1991, 1996). More recently, system dynamics modelers have incorporated facilitated, structured participatory methods in their modeling efforts (Van den Belt, 2004). In evaluation, visual causal diagrams are based on work done in theory-driven evaluation (Chen & Rossi, 1983) and logic modeling (Bickman, 1987; McLaughlin & Jordan, 1999; W. K. Kellogg Foundation, 2004). Ultimately, a well-articulated visual causal model provides the foundation for systematically and dynamically bridging research and practice.

Theory-Driven Evaluation

Theory-driven evaluation rests on the assumption that an intervention should be expressed in terms of its underlying causal relationships or program theory (Chen & Rossi, 1983; Weiss, 1997).

Program theory is both descriptive and prescriptive; it depicts the logical relationships within a program, outlining the forces at work between cause and effect and, most importantly, suggesting what should be done to optimize program results (Chen, 1990). Simply put, program theory describes in a straightforward way how a program is supposed to work (Bickman, 1987).

Many government agencies and nonprofit organizations have exhibited growing interest in theory-driven evaluation (Rogers, 2000); however, there have still been difficulties integrating theory-driven evaluation into practice (Weiss, 1997). Weiss (1997) points to complexities in constructing program theory, choosing which causal links to evaluate, and deciding between competing theories as some explanations for this implementation lag (Weiss, 1997, 2000). Despite the slow adoption of theory-driven evaluation in practice, it has proven to be a powerful tool for evaluators in many situations. It helps evaluators understand how a program works, what is needed for the program to work, and what can be done to provide evidence that the program does work (Sidani & Sechrest, 1999; Weiss, 2000). Previously, programs were often viewed as simply “black-boxes” that received inputs and produced outputs. Developing program theory allows evaluators to look inside the “black-box” and examine the mechanisms that lead to the desired outcomes (Bickman, 1987).

There are a variety of methods and tools for synthesizing and articulating program theory, including path analysis and causal modeling (Smith, 1990), the program theory matrix approach (Funnell, 2000), and concept mapping (Trochim, 1989; Trochim & Kane, 2005). One of the most recommended and widely used methods for program theory construction is the logic model approach.

Logic models

A logic model presents expected program performance in addressing targeted problems under certain conditions (Bickman, 1987; McLaughlin & Jordan, 1999; W. K. Kellogg Foundation, 2004). It can be understood as a “chain of reasoning” or set of “if-then” statements conditionally linking program components (W. K. Kellogg Foundation, 2004). The logic model is traditionally expressed in columnar format with abbreviated text in boxes connected by one-way arrows (McLaughlin & Jordan, 1999; Trochim & Cabrera, 2005).

Logic models ideally serve as planning, implementation, evaluation, and improvement devices (Kaplan & Garrett, 2005; W. K. Kellogg Foundation, 2004). In practice, the process of developing a logic model is often neglected and instead becomes an exercise in plugging items into a table (Renger & Hurley, 2006). In response to this common problem, Renger and Titcomb (2002) developed the three-step ATM approach (which stands for antecedent, target, measurement) to constructing the logic model that includes (a) identifying the antecedent conditions that underlie the rationale of the program; (b) targeting the antecedent conditions so that priority is given to those aspects that are most central to the agency mission or have the greatest support in the research literature; and (c) representing measurement in the model. One of the key strengths of the ATM approach is that it builds program theory on a solid research foundation. With a logical and well-supported causal chain, programs can justify long-term outcomes by demonstrating the existence of more short-term changes (Renger & Hurley, 2006).

Similar to the ATM approach, the SEP that we propose below emphasizes the need to relate the research literature to a clearly articulated program theory. There are several other similarities between the two approaches including (a) both emphasize the need for a causal diagram; (b) both highlight the need to target or limit the scope of an evaluation; and (c) both recognize that programs cannot typically measure long-term outcomes. Although both ATM and SEP ultimately result in a visual causal diagram, there are several fundamental differences in the process of developing and using the causal diagrams. The fundamental differences between ATM and SEP are (a) SEP is firmly grounded in theory including systems thinking, evolutionary theory, and developmental systems

theory and this underlies the processes and steps included in SEP. ATM does not appear to be theoretically grounded; (b) the goal of SEP is not to develop a single logic model (as is done in ATM) but rather a network of logic models that describe the larger system; (c) SEP begins by having practitioners articulate the specific program activities and subsequently identify the short-, medium-, and long-term outcomes. ATM begins with the outcome and identifies the problem or long-term goal and then works backward to outline the antecedents or preceding outcomes and activities. This is not a trivial distinction; (d) SEP can be implemented with an already existing program that is at any point in its lifecycle, whereas ATM is geared toward programs that are still in the planning phase; and (e) SEP views evaluation as part of a dynamically changing system where the scope of evaluation changes over time and builds from a focus on more proximal short-term outcomes to connecting with more general research that addresses long-term outcomes. ATM does not explicitly address the evolution of the scope of evaluation over time. While SEP uses a dynamic systems approach to linking evidence to evaluation, ATM does not focus on these dynamics and uses a more traditional static approach.

By applying a systems evaluation framework, SEP provides an approach that situates evaluators as the core facilitators of research–practice integration by including elements that assist practitioners with (a) critically thinking through program boundaries; (b) articulating a causal model; (c) visually linking the model with a research evidence base; and (d) continually assessing the evaluation plan over time to strategically build a case for the program’s success or use of relevant feedback to make program changes. Included in this approach are methods that aid in the clear and explicit expression of implicit theories (building the visual causal diagram), management of evaluation resources (pathway analysis and scope determination), and linkage of program theory and evaluation measures to a research evidence base (evidence mapping and measurement mapping). To illustrate the steps in this approach, we will refer to an example of a program evaluation plan developed for the Cornell Center for Materials Research (CCMR) to evaluate their module-based educational outreach efforts in STEM (science, technology, engineering, mathematics) education.

SEP: Building the Visual Causal Diagram

Before creating a visual causal diagram, a program boundary analysis should be conducted to determine what is “inside the program” and what is “outside the program.” Program boundary analysis is an exercise in language and terminology, which asks participants to clarify and make precise the statements they make about their program and may take for granted. Boundaries are artificial constructs created by humans and this analysis critically examines where they fall. Program names are simply labels given to a set of related activities and goals. The purpose of this analysis is to understand the meaning behind these labels and constructs (Cornell Office for Research on Evaluation, 2009). The result is a clear and precise program description that is agreed on by all the practitioners. At this point, they can discuss what activities are considered part of the program and subsequently the associated outputs and outcomes.

From Logic Models to Pathway Models

Similar to logic models, pathway models provide a conceptual framework for describing programs. However, while logic models rely on columnar representations that link whole sets of activities to sets of outcomes, pathway models make these connections more explicit and precise. Additionally, pathway models aid in the identification of orphaned activities and outcomes (activities and outcomes with no clear link to program logic) by helping to articulate a network of causal linkages, primary pathways, and nodes, thus, resulting in explanatory “throughlines” that explicitly connect

specific activities and outcomes. Each of the outcomes in a pathway model can be considered a node and the successive lines connecting activities and outcomes are pathways. Pathway models typically have multiple nodes and pathways. These throughlines, or pathways through the program's logic, subsequently aid in the development and expression of evaluation questions. Developing pathway models requires that the practitioners are not only able to describe the inputs, activities, and outcomes associated with their program, but that they also clearly make the implicit theories of programmatic logic explicit by articulating the connections between the various pieces.

Because many practitioners are not accustomed to making their implicit theories of programmatic logic explicit, we recommend beginning with a traditional columnar logic model. This enables practitioners to begin to think about the various components of their programs without having to detail immediately how they may be interrelated. Using the completed logic model as a starting point, practitioners can then develop their pathway model by drawing lines and directional arrows from a specific activity to specific short-, medium-, and long-term outcomes. Drawing these connections makes the still implicit causal logic expressed in logic models more explicit and unambiguous. Figure 1 provides an example of a completed pathway model for CCMR's module-based education outreach program.

Pathway Analysis and Determining the Scope

The resulting pathway model represents a broad overview of the key activities in a program and the expectations about the program's effectiveness and impact. However, in any given context, it is typically not possible to address every node and pathway through the model in the first cycle of an evaluation or even during the first few cycles. Therefore, a careful analysis of the pathway model is needed to determine the scope of an evaluation over any given period of time (e.g., quarterly, yearly).

There are several approaches that can guide the determination of the scope of an evaluation in any given year. The first approach uses the evaluation capacity present in the organization by building on already existing evaluation tools. The second approach focuses on maximizing evaluation resources by selecting nodes that affect multiple pathways. The third approach is perhaps the most common approach to determining the scope of an evaluation. One of the most typical constraints for evaluation scope is dictated by program funders and their requirements for evaluation. However, external mandates are only one way of approaching the scope of an evaluation. We will consider each of these approaches below.

We suggest using metaphors for describing these three approaches to determining evaluation scope. Metaphors not only help to describe the intentions of each of these approaches but also help practitioners and those new to evaluation to understand the use of carefully constructing and analyzing a program logic or pathway model. These approaches need not be used in isolation but can be combined for maximal impact.

Low-hanging fruit approach. When examining a pathway model, one of the first things to look for is whether there are any "easy" or obvious places to focus measurement. That is, we take into account our general capacity to accomplish evaluation and select opportunities that are both sensible and cost-effective. Begin by identifying the outcome or node that will be easiest to evaluate either because data are already being collected to address the outcome or because there are known preexisting measures for targeting that outcome. Then, identify the throughlines that connect this node to activities and other outcomes. This may be a good place to begin evaluation efforts, particularly in the first cycle of program evaluation. We term this technique the "low-hanging fruit" approach because the scope of the evaluation and the pathway of focus are determined by finding the easiest pathway to evaluate.

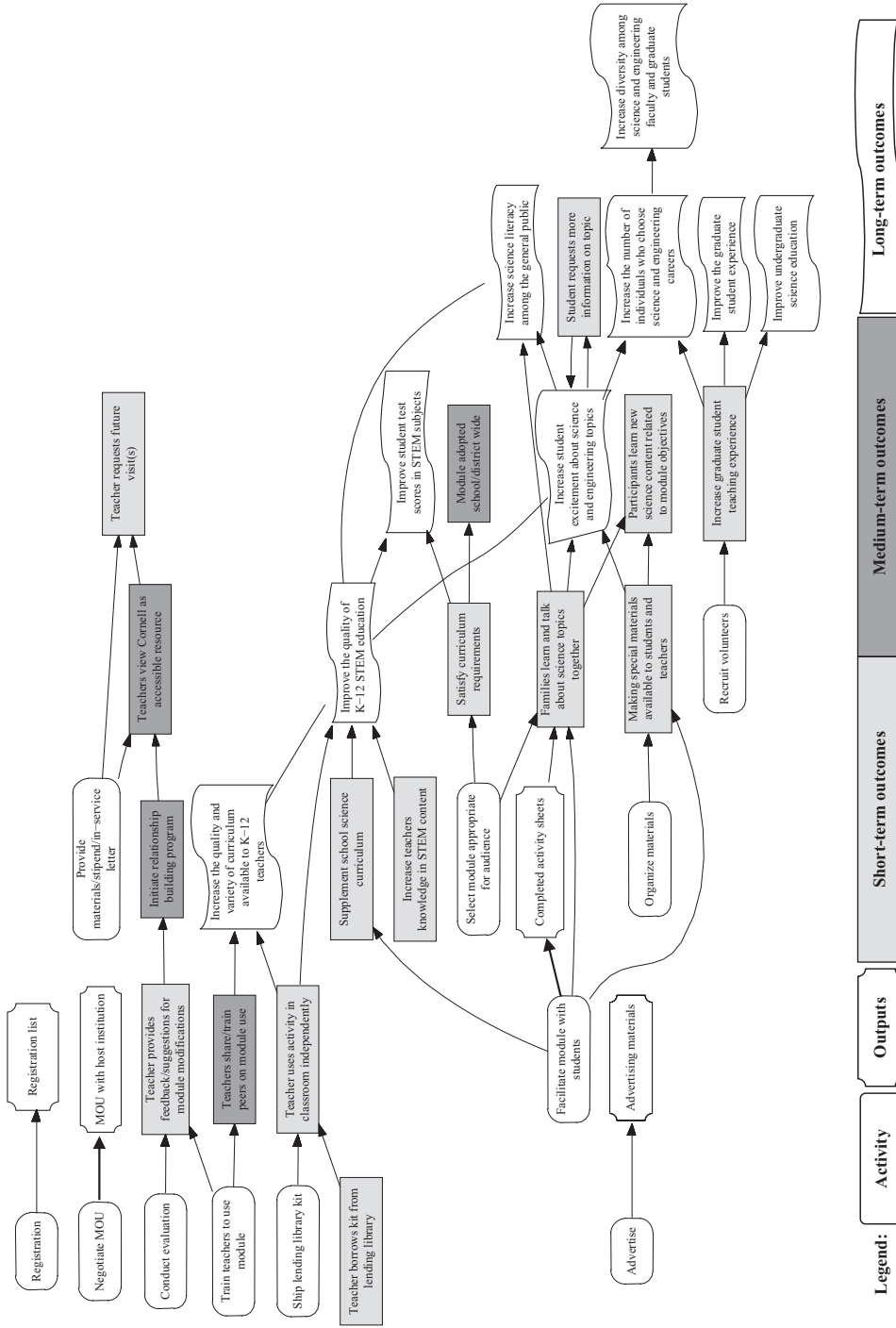


Figure 1. Pathway model for the Cornell Center for Materials Research (CCMR) module-based education program.

In the CCMR example, the low-hanging fruit is found at the short-term outcome node “teacher provides feedback/suggestions for module modifications” (Figure 2; see number 1). The organization is already collecting data on this outcome and has mechanisms in place for analyzing the data. In this approach, the throughline of focus will be on the relationship from the activity “conduct evaluation” all the way through the long-term outcome “teacher requests future visits” (Figure 2, see number 2a).

Grand central station approach. Another approach to determining evaluation scope and selecting specific pathways of focus can be accomplished by identifying the node (or outcome) that has the most pathways passing through it. A node with multiple pathways going into and passing out of it is termed a “hub.” Finding a “hub” helps to maximize resources. By measuring this outcome, the descriptive power of the evaluation is enhanced because many pathways can be addressed simultaneously. We term this technique the “grand central station” approach because the scope of the evaluation and pathways of focus are determined by identifying nodes that are key junctures for multiple pathways.

In the CCMR example, one hub is the short-term outcome “families learn and talk about science topics together” (Figure 2, see number 3). Multiple throughlines pass through this hub, and evaluating this outcome would consequently leverage evaluation resources. Throughlines begin at the activities “facilitate module with student” and “select module appropriate for audience.” They pass through the hub and on to several medium- and long-term outcomes including “increase science literacy among the general public” and “increase diversity among science and engineering faculty and graduate students” (Figure 2, see number 2b).

Do-or-die. The third approach is perhaps the most obvious and involves identifying the outcome/node on which the program is required to report and selecting the corresponding pathway.

In the CCMR example, the funder mandates that they evaluate the short-term outcome “making special materials available to students and teachers” and the long-term outcome “increase student excitement about science and engineering topics” (Figure 2, see number 4). Both these outcomes are predicated on the activity “organize materials” and result in the long-term outcomes “increase science literacy among the general public” and “increase diversity among science and engineering faculty and graduate students.”

Developing Evaluation Questions

Once the scope of the evaluation has been delimited and the pathways of interest have been identified, the specific evaluation questions for the current evaluation can be generated. In the CCMR example, we have decided to focus the next round of evaluation on the funder-mandated outcome “increase student excitement about science and engineering.” Therefore, the evaluation question is “does organizing the educational materials and making them available to students and teachers cause an increase in student excitement about science and engineering?” (Figure 2, see number 5). We also want to know whether this increase in excitement subsequently leads to an increase in the number of individuals who choose science and engineering careers that are long-term outcomes in the selected pathway.

SEP: Visually Linking Practice and Research

Once the causal visual diagram has been created, the evaluator’s next objective is to help practitioners make the connections between their program logic and a research evidence base. Most

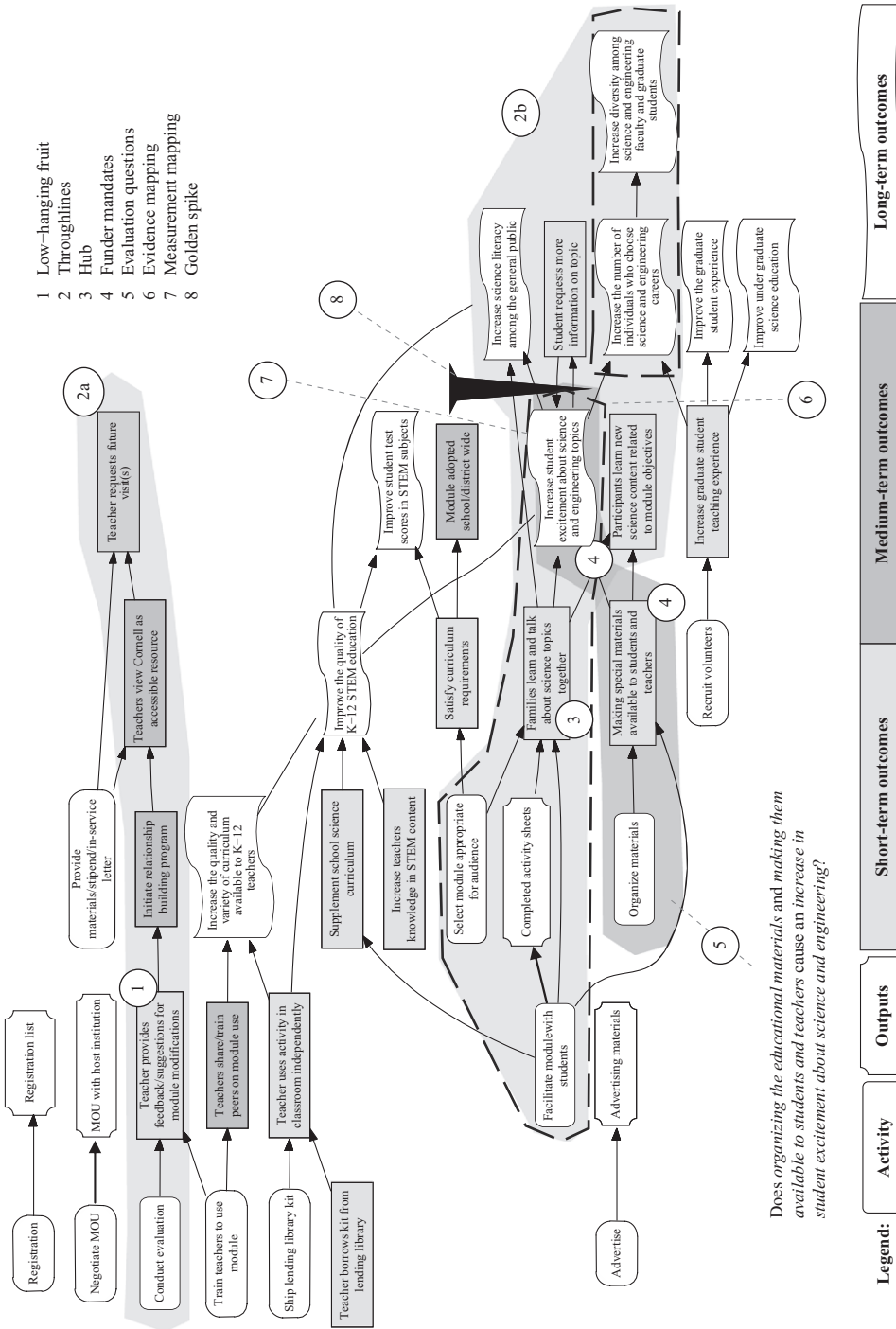


Figure 2. Pathway model for the CCMR module-based education program including identification of low-hanging fruit, throughlines, hub, funder mandates, evaluation questions, evidence mapping, measurement mapping, and the golden spike.

program evaluations cannot collect measures that will address all the outcomes of interest in the selected pathway. The long-term outcomes typically pose the greatest challenges for programs operating on tight budgets that do not have funds (or interest) to follow-up with program participants long after the end of the program. Articulating clear connections between the program logic and the research evidence helps to build the case for anticipated (but immeasurable) long-term outcomes. In addition, the research evidence base provides opportunities for cost and time savings by applying previously developed tools to the current evaluation questions.

Evidence mapping. Once the evaluation questions have been articulated, the research literature can be searched for evidence that might be relevant to the articulated causal logic. Ultimately, the goal is to link the program logic with an evidence base, a process we have termed “hanging” the literature on the causal relations that have been established.

Returning to the previous example, evidence for the relationship between student excitement about science and engineering and the pursuit of science and engineering careers might be identified in a research article that demonstrates that youth who show an interest in and excitement for science are more likely to pursue coursework in science in the future. We may also find a longitudinal study that began following youth in high school and measured student excitement about science and engineering and their subsequent pursuit of science-related careers (Figure 2, see number 6). Ideally, there will be multiple sources of evidence, all converging on the same finding, thus strengthening the connections between nodes in the pathway model. The more high-quality¹ evidence located that supports the causal relationships articulated in the pathway model, the stronger the case for the potential accomplishments of the program being evaluated, and the more successfully program managers will be able to link their locally evaluated programs with broader evidence of potential effectiveness.

Measurement mapping. In measurement mapping, we are looking for already existing measures that will aid in the measurement of the outcomes articulated in the pathway model and included in the evaluation scope. Ideally, we will find multiple measures, all of which converge on the same construct. The availability and feasibility of these measures may influence the decision of which outcomes will be measured. Whereas in evidence mapping the goal was to “hang” the evidence on the pathways between nodes, in measurement mapping, the goal is to “hang” measures on the nodes of the pathway model. Returning to our example, to address our primary outcome of interest, we might locate a validated survey for measuring student excitement about science and engineering, such as the Test of Science-Related Attitudes (TOSRA; Fraser, 1978; Figure 2, see number 7). This is an example where a survey worked for this context, but we encourage thinking broadly about measurement.

SEP and the Integration of Research and Practice: Locating the “Golden Spike”

Once the scope of the evaluation has been determined, the pathways of interest have been selected, and evidence and measurement mapping have been completed, we are ready to identify places where the evaluation efforts and the research literature meet, what we call the “golden spike.” The “golden spike” metaphor refers to the building of the transcontinental railroad, the joining of east and west that allowed for continuous transportation across the country. This reference suggests the coupling of two distinct realms, in the original instance that of the Central Pacific and Union Pacific railroads at Promontory Point, Utah, physically joining the east and west coasts of the United States. In this

context, it is meant to suggest the point at which research and practice connect, the network of pathways leading to that point, and the potential for dramatically extending our reach in evaluation through such a nexus.

The development of the visual causal model and the subsequent linkage with the research evidence base provides a framework for making the connections from short- and medium-term outcomes to long-term outcomes. The golden spike is literally a place that can be drawn on the visual causal map (Figure 2, see number 8). It is a place on the model where the evaluation results and the research evidence meet. The resulting connection between research and practice is compelling and useful for practitioners, researchers, and policy makers alike. Practitioners are able to address policy makers'/funders' requests for information on how their program is affecting long-term outcomes. Researchers benefit from the application of their research to real-world problems. Policy makers receive the data that they need to make decisions.

This linkage via visual causal models has particular implications for policy makers who are generally managing portfolios of programs or research that share common long-term outcomes. Policy makers are generally not concerned with any singular program evaluation or visual causal model but rather with the system of causal models that share common activities or outcomes. When considered in isolation, most programs or projects, even if perfectly implemented, have relatively minor effects when judged against some of the major systemic forces in peoples' lives. A small one-time science education experience is a relatively minor intervention compared to the other factors influencing middle school students. However, while any one program might be thought to induce relatively minor effects, it is more plausible that with the accumulation of multiple similar experiences, real change might be expected to occur. That is, although individual evaluations of small local interventions may show little or no effect on long-term outcomes, there may be a cumulative effect of multiple apparently ineffective individual programs. In other words, the true impact of a program may only be visible when viewed at a higher level of scale. Here, the pathway model, through the type of thinking implicit in the "golden spike" metaphor, can provide a compelling rationale for explaining how and why any given program can influence real change even if the only change they are able to demonstrate in their own evaluation is short-term and/or local. In this way, it helps to address both the practitioner's dilemma and the researcher's dilemma by connecting the local experience of the practitioner with the evidence of the researcher. In effect, it enables the practitioner's story about the effectiveness of a program to connect with the researcher's story about the scientific evidence enabling a continuous throughline that explains how a local program can have long-term impact.

The Dynamic Nature of Research–Practice Integration

A program will probably not reach the golden spike in its first evaluation cycle. Evaluation planning and implementation needs to be built over the life of the program until a clear link can be made between the research evidence and the activities and outcomes of the program. Rather than being a static event, evaluation is part of a dynamic process. The key is to look over time at how a program will build (via evaluation results) toward the golden spike. The primary purpose of the visual causal diagram is to provide a sense of where we need to push evaluation over time.

The Need for a Cyberinfrastructure: Imagining the Future

Although the approach outlined above can be implemented as a manual process that does not depend on any specific computerized platform, it is designed so that it can be enhanced throughout by using a cyberinfrastructure.

The key to the potential value of a cyberinfrastructure is in its use with systems of many programs. For example, it is reasonable to expect that many science education programs will share common key outcomes of interest such as “youth interest in science.” Although each of them can work independently to develop program logic and pathway models and might be successful in finding relevant research and research-based measures, a cyberinfrastructure that they use in common can potentially connect the programs to each other (so they can see and benefit from each other’s pathway models) and see and be seen by researchers (who can connect them with research and measurement tools). Funders and policy makers could view meta-summaries of projects across program areas, see where the programs are in their developmental lifecycles, and more effectively manage their portfolios of evaluations. Evaluators could in real time see their portfolio of projects and communicate with other users about their models and evaluation plans, thus enabling new models of virtual consultation. Field testing of these technologies is currently underway and holds the promise of extending the ideas suggested in this article to a potentially global audience of researchers and practitioners.

Conclusions

Making the long and difficult connections between research and practice, like making the cross-continental geographical connection of the transcontinental railroad, poses unique and daunting challenges. The problem is both a systemic one of integrating the worlds of research and practice and one that directly affects many program evaluations. Local program developers can feasibly conduct evaluations that demonstrate the effectiveness of their programs through short-term or perhaps even medium-term outcomes. However, few will have the resources or time to be able to conduct their own evaluations of the longer term outcomes and impacts of greatest interest to policy makers/funders and society as a whole.

The metaphor of the golden spike and the approaches associated with it offer a potential solution to this problem. If local program developers can push their successive program evaluations to a point where demonstrable outcomes can be linked to a line of research that demonstrates more systemic effects, then they can more plausibly argue from an evidence-based perspective that their programs contribute to such long-term outcomes. In our example, a local science education program that can demonstrate through evaluation that they have an effect on youth’s enthusiasm for science can join that throughline to the line of research that demonstrates that youth enthusiasm for science is associated with subsequent career choices later in life and more plausibly make the throughline case that their program contributes to subsequent career choices, even if they do not study them directly in the local setting.

The task for program developers is to push the scope of successive evaluations to golden spike points where they can connect with already determined research pathways. The task for researchers is to develop lines of research at higher levels of scope that link middle-term outcomes with longer term population effects and impacts and can be linked to by shorter term and more local efforts. The challenge for evaluation is to provide an environment and support systems to determine whether such connections are reasonable.

Of course, for many programs, there is the possibility that either no evidence exists to support the program model or evidence is identified that contradicts the proposed program model. Even in these cases, the approach offered here has considerable value. If there appears to be no relevant evidence, then the modeling effort will clearly show that none was identified. This then constitutes an important message to be sent from the realm of practice to that of research. Potential cyberinfrastructures should be used as a medium for alerting the research community about such research needs and gaps. If existing evidence is contrary to the theory of the program model, then it is especially important

that the practitioners embark on their work cognizant of that fact. Contradictory evidence would not on its own rule out the value of testing a program pathway model. There is always the potential that counter-indicating evidence could be generated. In the face of potentially conflicting evidence, the traditional options relevant in any research context apply: the evidence may be flawed, the theory may need to be revised, or both. In the modeling situations described in the article, the existence of contrary evidence should alert the practitioners to this fact and help them address these options. Currently, this step is seldom reached because thorough evidence searches seldom take place. In the approach offered here, such evidence searches are standard procedure.

In evaluation, as in life generally, a little planning can go a long way. Such evaluation planning should encompass the engagement of stakeholders, the development of a comprehensive model of what they think they are engaged in, and the articulation of a plan for evaluation that assesses the degree to which what they expected is what they observed and include contingencies for the unanticipated. Such evaluation planning is at the heart of efforts to integrate or connect the domains of practice and research; in effect, driving the golden spike to the point where these two domains intersect.

Well begun is half done.

Aristotle

Done!

Telegraph message from the “Golden Spike” Ceremony marking completion of the transcontinental railroad, May 10, 1869

Note

1. The question of what constitutes “high-quality” evidence is a topic that draws considerable debate and as such is beyond the scope of this article.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: supported in part by Grant Nos. 0535492 and 0814364 from the National Science Foundation.

References

- Antes, G. (1998). Evidence-based medicine. *Internist*, *39*, 899-908.
- Banerjee, S., & Dickinson, E. (1997). Evidence based health care in old age psychiatry. *International Journal of Psychiatry in Medicine*, *27*, 283-292.
- Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, *33*, 5-18.
- Brownson, R. C., Baker, E. A., Leet, T. L. & Gillespie, K. N. (Eds.). (2002). *Evidence-based public health*. New York: Oxford University Press.
- Busse, R., & Fleming, I. (1999). A critical look at cardiovascular translational research. *American Journal of Physiology*, *277*, H1655-H1660.
- Cabrera, D. (2006). *Systems thinking*. Unpublished PhD dissertation, Cornell University, Ithaca, NY.
- Cesario, A., Galetta, D., Russo, P., Margaritora, S., & Granone, P. (2003). The role of the surgeon in translational research. *Lancet*, *362*, 1082.

- Chen, H., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7, 283-302.
- Chen, H. T. (1990). Issues in constructing program theory. *New Directions for Program Evaluation*, 47, 7-18.
- Cornell Office for Research on Evaluation. (2009). *The evaluation facilitator's guide to: Systems evaluation protocol*. Ithaca, NY: Cornell Digital Print Services.
- Dougherty, D., & Conway, P. H. (2008). The "3T's" road map to transform US health care. *JAMA*, 299, 2319-2321.
- Durlak, J., & Dupre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350.
- Eitel, F., & Steiner, S. (1999). Evidence-based learning. *Medical Teacher*, 21, 506-512.
- Fraser, B. J. (1978). Development of a test of science related attitudes. *Science Education*, 62, 509-515.
- Funnell, S. C. (2000). Developing and using a program theory matrix for program evaluation and performance monitoring. *New Directions for Program Evaluation*, 87, 91-101.
- Galavotti, C., Kuhlmann, A., Kraft, J., Harford, N., & Petraglia, J. (2008). From innovation to implementation: The long and winding road. *American Journal of Community Psychology*, 41, 314-326.
- Gibbs, L. E. (2003). *Evidence-based practice for the helping professions*. Pacific Grove, CA: Thomson, Brooks-Cole.
- Green, L. (2006). Public health asks of systems science: To advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health*, 96, 406-409.
- Green, L. (2007). The prevention research centers as models of practice-based evidence: Two decades on. *American Journal of Preventive Medicine*, 33, S6-S8.
- Guerra, N., & Knox, L. (2008). How culture impacts the dissemination and implementation of innovation: A case study of the Families and Schools Together Program (FAST) for preventing violence with immigrant Latino youth. *American Journal of Community Psychology*, 41, 304-313.
- Horig, H., & Pullman, W. (2004). From bench to clinic and back: Perspective on the 1st IQPC translational research conference. *Journal of Translational Medicine*, 2, 44.
- Kaplan, S., & Garrett, K. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28, 167-172.
- Khoury, M. J., Gwinn, M., Yoon, P. W., Dowling, N., Moore, C. A., & Bradley, L. (2007). The continuum of translation research in genomic medicine: How can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genetics in Medicine*, 9, 665-674.
- Lesesne, C., Lewis, K., White, C., Green, D., Duffy, J., & Wandersman, A. (2008). Promoting science-based approaches to teen pregnancy prevention: Proactively engaging the three systems of the interactive systems framework. *American Journal of Community Psychology*, 41, 379-392.
- Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., et al. (2000). Measurement and psychotherapy—Evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, 177, 123-130.
- McDonald, P. W., & Viehbeck, S. (2007). From evidence-based practice making to practice-based evidence making: Creating communities of (research) and practice. *Health Promotion Practice*, 8, 140-144.
- McLaughlin, J., & Jordan, G. (1999). Logic models: A tool for telling your program's performance story. *Evaluating and Program Planning*, 22, 65-72.
- Mupparapu, M. (2002). Evidence-based dentistry. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics*, 94, 2-3.
- Olson, E. A. (1996). Evidence-based practice: A new approach to teaching the integration of research and practice in gerontology. *Educational Gerontology*, 22, 523-537.
- Renger, R., & Hurley, C. (2006). From theory to practice: Lessons learned in the application of the ATM approach to developing logic models. *Evaluation and Program Planning*, 29, 106-119.

- Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *American Journal of Evaluation*, 23, 493-503.
- Richardson, G. P. (1991). *Feedback thought in social science and systems theory*. Philadelphia: University of Pennsylvania Press.
- Richardson, G. P. (1996). *Modelling for management: simulation in support of systems thinking*. Aldershot: Dartmouth.
- Rogers, P. J. (2000). Causal models in program theory evaluation. *New Directions for Evaluation*, 87, 47-55.
- Rolleri, L., Wilson, M., Paluzzi, P., & Sedivy, V. (2008). Building capacity of state adolescent pregnancy prevention coalitions to implement science-based approaches. *American Journal of Community Psychology*, 41, 225-234.
- Sackett, D. L. (1997). Evidence-based medicine. *Seminars in Perinatology*, 21, 3-5.
- Saul, J., Duffy, J., Noonan, R., Lubell, K., Wandersman, A., Flaspohler, P., et al. (2008). Bridging science and practice in violence prevention: Addressing ten key challenges. *American Journal of Community Psychology*, 41, 197-205.
- Sidani, S., & Sechrest, L. (1999). Putting program theory into operation. *American Journal of Evaluation*, 20, 227-238.
- Smith, N. L. (1990). Using path analysis to develop and evaluate a program theory and impact. *New Directions for Program Evaluation*, 47, 53-57.
- Sung, N. S., Crowley, W. F. J., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., et al. (2003). Central challenges facing the national clinical research enterprise. *JAMA*, 289, 1278-1287.
- Ter Linde, J. J. M., & Samsom, M. (2004). Potential of genetic translational research in gastroenterology. *Scandinavian Journal of Gastroenterology*, 39, 38-44.
- Trochim, W. (1985). Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review*, 9, 575-604.
- Trochim, W. (1989). An introduction to concept mapping for planning and evaluation. *A Special Issue of Evaluation and Program Planning*, 12, 1-16.
- Trochim, W., & Cabrera, D. (2005). The complexity of concept mapping. *Emergence: Complexity and Organization*, 7, 11-22.
- Trochim, W., & Kane, M. (2005). Concept mapping: An introduction to structured conceptualization in health care. *International Journal for Quality in Health Care*, 17, 187-191.
- Trochim, W. M., Cabrera, D. A., Milstein, B., Gallagher, R. S., & Leischow, S. J. (2006). Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health*, 96, 538-546.
- Van den Belt, M. (2004). *Mediated modeling: A system dynamics approach to environmental consensus building*. Washington, DC: Island Press.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., et al. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41, 171-181.
- Weiss, C. (1997). *Evaluation*. Upper Saddle River, NJ: Prentice Hall.
- Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation*, 87, 35-45.
- Westfall, J. M., Mold, J., & Fagnan, L. (2007). Practice-based research—"Blue Highways" on the NIH road-map. *JAMA*, 297, 403-406.
- Williams, B., & Imam, I. (Eds.). (2006). *Systems concepts in evaluation: An expert anthology*. Point Reyes CA: Edge Press.
- W.K. Kellogg Foundation (2004). *W.K. Kellogg Foundation logic model development guide*. Battle Creek, MI: W.K. Kellogg Foundation.
- Woolf, S. H. (2008). The meaning of translational research and why it matters. *JAMA*, 299, 211-213.
- Young, G. (1999). A piece of the main: Parts and wholes in ecology and human ecology. *Advances in Human Ecology*, 8, 1-31.