Scientific accompaniment: a new model for integrating program development, evidence and evaluation

Patricia Lannen and Lisa Jones

Abstract

Purpose – *Calls for the development and dissemination of evidence-based programs to support children and families have been increasing for decades, but progress has been slow. This paper aims to argue that a singular focus on evaluation has limited the ways in which science and research is incorporated into program development, and advocate instead for the use of a new concept, "scientific accompaniment," to expand and guide program development and testing.*

Design/methodology/approach – *A heuristic is provided to guide research–practice teams in assessing the program's developmental stage and level of evidence.*

Findings – *In an idealized pathway, scientific accompaniment begins early in program development, with ongoing input from both practitioners and researchers, resulting in programs that are both effective and scalable. The heuristic also provides guidance for how to "catch up" on evidence when program development and science utilization are out of sync.*

Originality/value – *While implementation models provide ideas on improving the use of evidence-based practices, social service programs suffer from a significant lack of research and evaluation. Evaluation resources are typically not used by social service program developers and collaboration with researchers happens late in program development, if at all. There are few resources or models that encourage and guide the use of science and evaluation across program development.*

Keywords *Evaluation, Research–practice gap, Evidence-based interventions, Prevention, Program maturity, Children*

Paper type *Viewpoint*

List of abbreviations

CDC = Center for Disease Control; and
TOC = Theory of Change.

1. Introduction

The research practice gap has been extensively described and debated in health, education and social service fields (Backer *et al.*, 1995; Chambers, 2012; Flaspohler *et al.*, 2012a, 2012b; Glasgow and Emmons, 2007; Green *et al.*, 2012; Hallfors and Godette, 2002; Morrissey *et al.*, 1997). Despite numerous books, manuals, websites and other publications on evaluation and a long-standing call for evidence-based interventions in many fields (Eagle *et al.*, 2003; Lyles *et al.*, 2006; Nathan and Gorman, 2015; Truman *et al.*, 2000; Zaza *et al.*, 2005), progress in narrowing the research practice gap has been slow (Fagan *et al.*, 2019; Gambrill, 2016; Neuhoff *et al.*, 2022; Wathen and MacMillan, 2018). The goal of better integrated research and program implementation are hampered by several problems including:

Patricia Lannen is based at the Marie Meierhofer Children's Institute, University of Zurich, Zurich, Switzerland. Lisa Jones is based at the Crimes Against Children Research Lab, University of New Hampshire, Durham, New Hampshire, USA.

- limited backgrounds in research and science for most practitioners;

- pressure on practitioners to incorporate "evaluation" and demonstrate their program is evidence-based, but without much guidance on what kind of evaluation makes sense at which stage or how to connect with a good partner;

- research-practice partnerships that often begin late in the program life cycle; and

- a mismatch between researcher skills and interests and practitioner needs.

To better understand how these problems typically arise, it is helpful to consider how programs often develop. There are two different approaches for developing programmatic interventions that have been described: the *research to practice* model, and the *community-centered model* (Wandersman *et al.*, 2008). *Research to practice* interventions generally include research from the onset (Wandersman *et al.*, 2008) and follow predefined stages for program development (Mercy *et al.*, 1993; Mrazek and Haggerty, 1994). There have been some great successes of such models such as Nurse Family Partnerships (Olds, 2006), Perry Preschool Program (Heckman *et al.*, 2010) and the Incredible Years (Menting *et al.*, 2013). However, many times, interventions developed and tested by researchers do not transfer easily to practice (Glasgow and Emmons, 2007; Green *et al.*, 2009; Wandersman *et al.*, 2008) and are difficult to implement in many parts of the world due to resource constraints, logistics or a lack of implementation expertise by researchers (Miller and Shinn, 2005; Richardson, 2009; Wandersman *et al.*, 2008; Ward *et al.*, 2014). Others have criticized that programs developed by researchers are rigid, ignore participant or user values and practices and might not be adaptable to individual needs of the participants (Mullen and Streiner, 2004). Hence, overall interventions that have been evaluated and found to be most effective in prevention research are not necessarily the ones most widely implemented (Ringwalt *et al.*, 2002; Wandersman and Florin, 2003).

Community-centered models, on the other hand, are more commonly implemented and are developed primarily by individuals with strong advocacy or practice expertise (Wandersman *et al.*, 2008). However, they rarely have access to research or science expertise and their limited time and resources focus on developing and implementing the program itself (Green *et al.*, 2009; Jones, 2014; Wandersman *et al.*, 2008). For these models, research–practice partnerships are critical to strengthening programmatic evidence and improving the knowledge of what works (Kelly, 2012). In our experience though, such partnerships, if they are undertaken at all, typically happen late in program development, well after the program has been implemented widely or even scaled. The partnerships also sometimes occur around a specific pressure to include evaluation by outside sources, such as by a particular funding requirement.

As researchers who have consulted with nonprofit organizations working to improve the lives of children and families for decades, and in the case of one of us, worked for a major private international funder that commissioned research and evaluations, we have seen these challenges regularly interfere in efforts to build more effective programs for children and families. We believe that a new orientation is needed that includes several key features:

- novel, explicit and realistic approaches to guiding research–practice partnerships;

- models in which science and research is incorporated early in program development; and

- a roadmap for program developers (and evaluators) that provides guidance on the kind of evaluation or science use that makes sense given a program's stage of development.

To encapsulate this approach, we suggest the use of term "scientific accompaniment," borrowing from the German term *wissenschaftliche Begleitung*, which is sometimes translated as "concomitant research," (Bär, 2013). We believe that the field's reliance on

evaluation may limit the ways that program developers use science and research in their work. We hope that the use of the term scientific accompaniment will instead move researchers and practitioners to consider a process that builds research into early stages of development, increasing the rigor of programmatic evidence step by step, throughout the entire program life cycle. The term scientific accompaniment also promotes a co-constructive process, in which scientific knowledge and available evidence is integrated with expertise from practice and advocacy provided by the practitioner to build both effective and scalable programming. This article discusses the concept of scientific accompaniment, providing a heuristic for how research–practice teams can best collaborate based on program maturity. It first outlines and defines different stages of program development as well as different levels of evidence for programs. It then outlines the types of scenarios that research–practice partnerships may encounter depending on a program's stage of development and level of evidence. Finally, the paper provides suggestions for research–practice teams on how to best proceed to strengthen a program's evidence.

2. A heuristic for determining program maturity

The term program maturity has traditionally been used synonymously with a program's stage of development (Baker and Perkins, 1984; Milstein and Wetterhall, 1999). However, defining a program as mature should indicate that it has been developed, tested and grown over time along with *accompanying science and evaluation support*. Using this more multi-faceted definition, we suggest that research–practice teams assess the stage of program maturity according to a heuristic that combines the program's stage of development with its level of evidence support.

2.1 Stage of program development

There are numerous frameworks that describe stages of program development and implementation (Birken *et al.*, 2017; Meyers *et al.*, 2012; Tabak *et al.*, 2012). In their seminal work, the Institute of Medicine published a five-stage model to develop successful interventions (Mrazek and Haggerty, 1994). Another pioneering model is the Center for Disease Control model developed by Mercy and colleagues (Mercy *et al.*, 1993), which outlines four stages of program development: defining the problem, identifying risk factors, developing and testing interventions, ensuring widespread use. For the purpose of engaging in scientific accompaniment, we define a four-stage model of program development.

Concept stage. This stage describes the phase in which a community need is identified, and a solution to the problem in the form of an intervention emerges and is planned. Program components are developed and the logistics of implementation are considered.

Pilot stage. The pilot stage of a project is an initial small-scale implementation that is used to test whether the project idea is viable with a small number of beneficiaries. It enables the organization to assess and manage risks of a new idea and identify any need for improvements before substantial resources are invested.

Implementation stage. The implementation stage is the phase in which the project is actually executed with a targeted number of beneficiaries. It has been suggested that it takes two to four years to solidify effective implementation of a program (Fixsen *et al.*, 2009).

Scale up stage. This stage describes a program that was designed for one setting and is now being more widely implemented in other locations with the same or very similar settings (Aarons *et al.*, 2017).

2.2 Levels of evidence

In considering the evidence support for program impact, we use a broad definition of evidence that includes research or evidence for a program's approach or components, evidence for the need for a program, data on implementation feasibility and evidence of a

program's impact or effectiveness. While there have been some attempts to label different levels of evidence for an intervention (Dekkers, 2018; Geher, 2017), no established terminology exists. In the current paper, we define three levels of evidence:

Not supported by evidence. We use the category "not supported by evidence" to refer to a program that:

- has not systematically gathered information on the problem, its risk factors, effective solutions for a problem or proven mechanisms of change in their conceptual framework;

- has not used established research on related social problems to develop a Theory of Change (TOC) and inform program components (Darling *et al.*, 2016; De Silva *et al.*, 2014; Jones, 2014; Valters, 2015; Valters *et al.*, 2016; Weiss, 2011); and

- has not undergone a process or outcome evaluation.

Evidence-informed. We use the term "evidence-informed" to refer to programs that, while they may not yet have used outcome evaluation to confirm program efficacy, the program design and implementation have been designed using available data. Specifically, we consider this category as having two levels: first, the intervention has been embedded in a TOC that considers existing research results. Second, the program uses developmental or process evaluation to answer questions relevant to practitioners that arise during the course of program design, implementation and refinement (Peters *et al.*, 2013). While "good" interventions can be badly implemented, poor interventions can equally be implemented successfully. Having theoretically sound programs does not, in itself, ensure successful implementation and/or effectiveness (Moir, 2018).

Evidence-based. We refer to evidence-based programmatic interventions as those that have undergone some kind of formal evaluation with evidence of positive impact on at least some key outcomes. There are a range of evaluation methodologies with different levels of rigor that have been well described in multiple texts. (Rossi *et al.*, 2019; Wholey *et al.*, 2010). There are also variations and differences of opinion on what level of evidence is needed to consider a program evidence-based (Mihalic and Elliott, 2015). In our view, the use of this term should refer less to an end result, than to a process of building evidence for program impact with increasing rigor as the program matures. Smaller outcome evaluations with less rigorous designs (e.g. pre-post designs) can help tweak program components earlier in development. Then, increasingly rigorous methodologies (e.g. randomized controlled trials) are needed before a program is scaled.

2.3 Program maturity

When a research–practice partnership is initiated, the first goal of the team should be to assess the current level of program maturity by defining: the program's stage of development; and the level of available evidence for the program. We have developed a heuristic depicted in Table 1 to outline the different scenarios that researchers can encounter at the outset of a partnership with practitioners based on a combination of these two factors. Based on the assessment and the resulting scenario, the team can systematically assess the key next steps for collaboration and how to best approach moving the evidence forward to "catch up" on missed steps.

Some of the scenarios represent more ideal situations for research–practice partners than others [1]. In the best case, a research–practice partnership will be initiated in Scenario 2, then move through Scenarios 8, 9, 15 and 20 with the following recommendations:

- A program should not move into a next phase along this path until the suggested level of evidence has been reached. For example, program should not be scaled up until rigorous evaluation (with control groups) indicate that the program is effective.

	Stages of program development			
Table 1 A heuristic for determining program maturity				
Evidence support for program	Concept phase	Pilot phase	Implementation phase	Scale Up phase
Not supported by evidence	1	6	11	16
Evidence-informed				
Level 1: TOC	2	7	12	17
Level 2: developmental or process evaluation	3	8	13	18
Evidence-based				
Level 1: pre-post outcome evaluation	4	9	14	19
Level 2: outcome evaluation with control group	5	10	15	20
Note: TOC = Theory of change				

- A certain level of evidence should not be generated until all the steps have been "caught up." For example, a program should hold off from rigorous evaluation until an evidence-informed TOC exists and the project has been evaluated for process and some basic understanding of effectiveness has been established through pre-post evaluation.

The remaining scenarios represent situations in which a program's stage of development and level of evidence may be out of sync, which, in our experience, is not uncommon. For those scenarios, we use the heuristic to provide guidance to research–practice teams on how to work together in these conditions.

3. Scientific accompaniment: a roadmap for collaboration between researcher and practitioner

The following sections discuss the process of scientific accompaniment for each of the four stages of program development (concept phase, pilot phase, implementation phase and scale-up phase) using the heuristic presented in Table 1. We provide guidance, in particular when a research-practice collaboration is initiated at all the different stages of program maturity. In each section, we outline the opportunities and risks that exist within the different scenarios and identify strategies for making sure that a program's level of evidence and stage of development are in sync.

3.1 Concept phase

The integration of science and research and practitioner expertise during the concept development phase for social programs is critical, but evaluation and implementation literature have typically not focused much attention here. Despite calls to link research and practice in the process of intervention design and testing, it is rare for research–practice partnerships to occur at this phase (Glasgow and Emmons, 2007; Green et al., 2009; Miller and Shinn, 2005; Wandersman, 2003). Our program maturity heuristic defines Scenario 2 as an ideal place for the scientific accompaniment process to begin. Research–practice partnerships occurring at this stage of program development increase the chance that programs will have positive outcomes. It is also an effective way to build long-term partnerships: researchers and practitioners build a common language and a common understanding of the program as it develops. Early work together can also facilitate efforts to define outcomes that might be measured in future outcome evaluations and make preparations that will facilitate outcome evaluation work as the program develops.

With a research-practice partnership during this phase, practitioners' expertise can be connected with existing evidence to make sure that the program design and implementation plan is evidence-informed. This process involves two tasks. The first task is

to clarify the community or population's need for the program, such as through a needs assessment (Soriano, 2013). Collaboration with researchers can provide data to confirm the practitioner's experience-based impression about what kinds of services are needed for a given target population.

The second task is to define the program components and connect them through an evidence-informed TOC. In a strong TOC, program elements and outcomes are well-defined, and the assumptions connecting them are backed up with evidence (Darling *et al.*, 2016; De Silva *et al.*, 2014; Jones, 2014; Valters, 2015; Valters *et al.*, 2016; Weiss, 2011). While many programs construct a TOC (or a logic model or program theory), often the assumptions underlying the TOC are not clearly supported by evidence. Prior evaluation and research efforts help practitioners design components or mechanisms of change so that, from the start, the program has the best possible chance of being effective (Leijten *et al.*, 2018; Melendez-Torres *et al.*, 2019). Even if a program is being developed to tackle a relatively new problem area, there is likely evaluation research on programs addressing related areas, or research on risk and protective factors, that can be used to support program development.

3.2 Pilot phase

In a program's pilot phase, the use of science and research allows practitioners to test feasibility, address challenges in program delivery, establish procedures, collect preliminary data on participants and make changes to the model prior to expanding delivery to full implementation (Wiseman *et al.*, 2007). The aim for this phase is to end up with a fully evidence-informed program design, ready for implementation and outcome evaluation. When teaming up with practitioners, researchers might encounter a program in a pilot phase that has not drawn from existing research or data yet (Scenario 6). In this case, because of the early stage of development, there is an opportunity to review the TOC and identify areas where research supports the program logic and where it does not. The research–practice team are then able to jointly strengthen the TOC with available evidence and refine program elements if work on the TOC reveals that changes are needed to the model or approach.

Once the team is working with an evidence-informed TOC (Scenario 7 in the heuristic), the key focus of the research–practice partnership needs to be on systematically collecting data on how the pilot and early implementation is going through the use of a formative or process evaluation (Cohen *et al.*, 2000; Crowther and Lancaster, 2012). The goal of such an evaluation is to collect information that can be fed back into program implementation, and there are extensive resources to guide this process (Patton, 1994; Patton *et al.*, 2016). It answers questions such as:

- *Q1.* How well is the program being implemented?
- *Q2.* Is it implemented as planned?
- *Q3.* How well is the target population being reached?
- *Q4.* What challenges have been revealed?
- *Q5.* What possible solutions have been tested?
- *Q6.* What elements have proved useful/popular by the recipients, and which ones have not?

Typically, formative evaluation is a term used early in program development stages (as in a pilot stage), while the term process evaluation can refer to data collected at any point in a program's development to ensure that implementation is proceeding as intended (Wholey *et al.*, 2010). Formative evaluations have been shown to be crucial in the process to

strengthen an intervention and get it ready to be evaluated for impact more rigorously (Devries *et al.*, 2021; Lachman *et al.*, 2020; Madrid *et al.*, 2020)

It may also be informative to collect preliminary data on pilot effectiveness in a pre-post design, depending on the number of individuals participating in an intervention and the nature of the pilot (Heuristic scenario 9). Some organizations may want to verify likelihood of effectiveness in a larger-scale pilot before expanding to an implementation phase. Other organizations may do developmental evaluation with a small pilot, and then focus on outcome evaluation as part of the implementation phase.

There may be programs interested in engaging in more rigorous outcome evaluation (e.g. involving control groups) before moving to the implementation stage (Heuristic scenario 10) (Arain *et al.*, 2010). This may be the case, for example, with programs developed in controlled academic settings who want to use an efficacy trial as part of early program development (Flay, 1986; Glasgow *et al.*, 2003). Sometimes, the practitioners or researcher may be impatient to move to outcome evaluation. Practitioners may be eager to claim that they have an "evidence-based program"; researchers may have an incentive to lead rigorous evaluations with a high chance of being published. Both may have heard that highly rigorous designs, such as randomized controlled trials are gold-standard evaluations and are keen to move to that level of recognition. However, conducting rigorous outcome evaluation during the pilot stage bears risks for a program. Even efficacy trials require that programs have achieved some finalized and ready stage of development and stable implementation. Additionally, conducting rigorous evaluation in the midst of program development might produce results that reflect delivery issues still being sorted out versus the impact of program elements.

3.3 Implementation phase

The implementation phase is defined by programs that are more established and are being delivered as part of normal organizational procedures to a larger group of individuals. During this phase, the program should aim to move to a place where program efficacy is established as the program builds, and prior to any scaling. The implementation phase is typically variable in length, in which a program moves from early implementation to sustained and long-standing community programs. Some programs remain here as locally implemented programs without moving to expanded, scaled implementation. It is our experience that this is the most common phase during which a research–practice partnership is initiated. For example, a researcher may be brought on by a practitioner who wants to conduct evaluations to see if the program is working. This can happen early in implementation, but our experience is that often research–practice partnerships occur after a program has been well-established, and practitioners have become interested in documenting impact.

There are challenges that will need to be addressed if there has been little prior attention to the TOC (Scenario 11) or developmental evaluation work (Scenario 12). However, research–practice teams can work to address missed steps. There is some possibility that addressing the TOC at this stage might uncover significant gaps in program logic. In this case, moving too quickly to a rigorous evaluation would potentially waste resources, and instead, an adaptation of the program might be warranted before moving into a formal evaluation. It may be difficult for programs to consider changing program elements, particularly if the research–practice partnership is new and still building trust. However, it is better to do a correction now, before resources are spent on an outcome evaluation with negative outcomes and certainly before the program is scaled up. Similarly, research-practice teams should make sure that process evaluation work precedes outcome evaluation, so that evaluation work is not being conducted on programs that are not being delivered fully or with a basic level of fidelity to the design (Scenario 13).

Outcome evaluation is critical to conduct during the implementation phase with several factors influencing the decision of the research–practice team about the rigor of outcome evaluation that should be conducted (Scenarios 14 and 15). It might be wise to collect pre-post data only in a first step to solidify hypotheses related to outcomes of interest, before conducting a study producing more confidence in effectiveness with a more rigorous design (Habicht *et al.*, 1999). Evaluation research should build on prior work, becoming more rigorous and more independent over time. Prior to scaling a program, it is ideal for rigorous and independent evaluation to confirm program impact with at least some key outcomes. Randomized control trials are considered the gold standard of outcome evaluation, although there are some logistical as well as resource considerations to be taken into account (Sanson-Fisher *et al.*, 2007; West *et al.*, 2008). Still, it is important to make sure that implementation is solid before engaging in rigorous impact evaluation. Many of the risks and challenges outlined under the pilot phase still apply. In addition, practitioners may get frustrated to try to incorporate large-scale evaluation as the program still works on building implementation objectives. (Carroll *et al.*, 2007). A program might also fail to show effects because the number of participants is too small. In addition, conducting a rigorous evaluation early on might produce results indicating effectiveness in a very specific setting and render the program inflexible for other contexts (Glasgow *et al.*, 2003).

Choosing the appropriate research design depending on the maturity of the program is key to mitigating risks while gradually strengthening evidence. At the same time, and depending on their training and focus, researchers may be inclined and more comfortable to conduct a certain type of evaluation, for example, process evaluation for more qualitatively trained researchers or outcome evaluation for more quantitatively trained researchers. Practitioners might also have a preference for one type of evaluation or another, depending on their understanding of evaluation or their ultimate purpose as opposed to choosing the design based on the research question, as has been strongly suggested (Peters *et al.*, 2013). Often, the necessary researcher skills include qualitative as well as quantitative elements and a mixed approach is generally useful; hence, it is important to make sure that the research questions and methodology match researcher skills.

3.4 Scale-up phase

Many programs with successful implementation become interested in moving implementation to other communities or even large geographical regions. As a general recommendation, a program should only be scaled up once sufficient evidence of effectiveness is available. The scale up process brings many new challenges of implementation and also additional questions about effectiveness (Forum on Promoting Children's Cognitive, Affective, and Behavioral Health, Board on Children, Youth, and Families, Institute of Medicine, and National Research Council, 2014). While implementation may be very successful in one setting or community, there may be unexpected implementation challenges in new ones. And even strong outcome evaluation in one setting does not guarantee the same level of impact in a new setting (Olweus and Limber, 2010). Hence, it is important to continue process and outcome evaluation during scale up process for quality assurance. Aarons *et al.* (2017) argue that when implementing interventions in a moderately different setting or with a different population, it can sometimes "borrow" strength from evidence of impact in a prior effectiveness trial, but argue that some new empirical evidence is often necessary to retain evidentiary status.

However, many programs are scaled before any evidence of effectiveness has been established at all, with dire consequences. A famous example is Drug Abuse Resistance Education (D.A.R.E.): this program designed to prevent drug use was widely disseminated, administered in 70% of US school districts in 1996 (Rosenbaum and Hanson, 1998). Once evaluated, however, it was deemed ineffective (West and O'Neal, 2004) and even potentially harmful (Lilienfeld, 2007). Should a research–practice partnership be initiated with a program

that has already been scaled up, the different stages of strengthening the evidence for a program provided in the heuristic should occur before moving to rigorous outcome evaluations (Heuristic scenarios 18, 19 and 20). The broader the scale of implementation, the more difficult it may be to adjust the program according to the work done in these scenarios. However, building in scientific accompaniment is still of critical value and can make a difference. Based on the findings from the evaluations, the D.A.R.E. program completely revised their curriculum with improved evidence of effectiveness in multiple, rigorous and controlled studies (Hecht *et al.*, 2006; Marsiglia *et al.*, 2011).

4. Conclusions

Aiming toward a model of "scientific accompaniment" will expand the orientation of both practitioners and researchers beyond just outcome evaluation. It encourages the use of research-practice partnerships from the beginning, throughout the program's life cycle and provides a roadmap for researcher–practice collaboration. Through the use of a heuristic, it outlines scenarios that define idealized pathways for syncing program development and the use of science, as well as how teams can "catch up" on evidence, should evidence and program development end up out of sync.

The model is particularly useful because a research–practice partnership can be initiated at any stage of program development, and at any level of pre-existing evidence. It frees the collaboration from a key stumbling block in our experience, namely, to implement the type of evaluation a researcher might be most familiar with or the evaluation the practitioner is envisioning for the program. Instead, the heuristic guides the partnership to carefully match the "science intervention" with program maturity, providing a useful "rule of thumb" on how to best proceed at this point in time.

The model makes sure to use available evidence to strengthen programming as well as systematically harvesting expertise from practitioners, responding to the call that efforts to close the gap should include both researcher and practitioner perspective (Morrissey *et al.*, 1997; Wandersman, 2003). The model provides a realistic pathway to build increasing confidence that the program design is strong, that implementation is needed and feasible, and the that program is having the anticipated benefit in the recipients and perhaps community. Evaluation theory is increasingly emphasizing the importance of value-engaged approaches or making sure that evaluation incorporates stakeholder values (Hall *et al.*, 2012). Realist evaluation theory emphasizes the importance of evaluation as an iterative process working to understand the nuance of how a particular program might work with a given population and setting and why (Jagosh *et al.*, 2015; Marchal *et al.*, 2012). A model of scientific accompaniment versus evaluation allows more easily for incorporation of these theoretical perspectives, which developed out of concern that a traditional, narrow approach to evaluation does not account for the complexity of the real-world contexts in which programs seek to make change.

While this model presents a structured approach to establishing the evidence-base of a given intervention and formulates a "roadmap," there may be additional issues that need to be taken into consideration during implementation. This may include challenges to receive funding for the initial phases of implementing the scientific accompaniment model, or challenges of researchers remaining involved in all phases of the research. However, there are a number of examples where this has been successfully done. As models of scientific accompaniment expand, the benefits are likely to become even more apparent to funders and other stakeholders in evidence-based practice.

Note

1. For the concept phase, only Scenarios 1 and 2 are relevant, as the pre-condition for both process and impact evaluation is that the program is implemented at least with a small number of beneficiaries and not only exists as a concept.

References

Aarons, G.A., Sklar, M., Mustanski, B., Benbow, N. and Brown, C.H. (2017), "'Scaling-out' evidence-based interventions to new populations or new health care delivery systems", *Implementation Science*, Vol. 12 No. 1, p. 111, doi: 10.1186/s13012-017-0640-6.

Arain, M., Campbell, M.J., Cooper, C.L. and Lancaster, G.A. (2010), "What is a pilot or feasibility study? A review of current practice and editorial policy", *BMC Medical Research Methodology*, Vol. 10 No. 1, pp. 1-7.

Backer, T.E., David, S.L. and Soucy, G. (1995), "Reviewing the behavioral science knowledge base on technology transfer", *NIDA Research Monograph*, Vol. 155, pp. 1-20.

Baker, F. and Perkins, D.V. (1984), "Program maturity and cost analysis in the evaluation of primary prevention programs", *Journal of Community Psychology*, Vol. 12 No. 1, pp. 31-42, available at: https://doi.org/10.1002/1520-6629(198401)12:1<31::AID-JCOP2290120105>3.0.CO;2-7

Bär, G. (2013), "Wissenschaftliche Begleitung, formative evaluation und partizipative forschung", *Prävention Und Gesundheitsförderung*, Vol. 8 No. 3, pp. 155-162, doi: 10.1007/s11553-013-0397-y.

Birken, S.A., Powell, B.J., Shea, C.M., Haines, E.R., Alexis Kirk, M., Leeman, J., Rohweder, C., Damschroder, L. and Presseau, J. (2017), "Criteria for selecting implementation science theories and frameworks: results from an international survey", *Implementation Science*, Vol. 12 No. 1, p. 124, doi: 10.1186/s13012-017-0656-y.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S. (2007), "A conceptual framework for implementation fidelity", *Implementation Science*, Vol. 2 No. 1, p. 40, doi: 10.1186/1748-5908-2-40.

Chambers, D.A. (2012), "The interactive systems framework for dissemination and implementation: enhancing the opportunity for implementation science", *American Journal of Community Psychology*, Vol. 50 Nos 3/4, pp. 282-284, doi: 10.1007/s10464-012-9528-4.

Cohen, L., Manion, L. and Morrison, K. (2000), *Research Methods in Education*, 5th ed., Routledge, London, doi: 10.4324/9780203224342.

Crowther, D. and Lancaster, G. (2012), *Research Methods*, Routledge, Waltham Abbey.

Darling, M., Guber, H., Smith, J. and Stiles, J. (2016), "Emergent learning: a framework for whole-system strategy, learning, and adaptation", *The Foundation Review*, Vol. 8 No. 1, doi: 10.9707/1944-5660.1284.

De Silva, M.J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C. and Patel, V. (2014), "Theory of change: a theory-driven approach to enhance the medical research council's framework for complex interventions", *Trials*, Vol. 15 No. 1, p. 267, doi: 10.1186/1745-6215-15-267.

Dekkers, H. (2018), "Science-based, research-based, evidence-based: what's the difference?", *Dynaread*, available at: www.dynaread.com/science-based-research-based-evidence-based

Devries, K., Balliet, M., Thornhill, K., Knight, L., Procureur, F., N'Djoré, Y.A.B., N'Guessan, D.G.F., Merrill, K.G., Dally, M., Allen, E., Hossain, M., Cislaghi, B., Tanton, C. and Quintero, L. (2021), "Can the 'learn in peace, educate without violence' intervention in cote d'Ivoire reduce teacher violence? Development of a theory of change and formative evaluation results", *BMJ Open*, Vol. 11 No. 11, p. e044645, doi: 10.1136/bmjopen-2020-044645.

Eagle, K.A., Garson, A.J., Beller, G.A. and Sennett, C. (2003), "Closing the gap between science and practice: the need for professional leadership", *Health Affairs*, Vol. 22 No. 2, pp. 196-201, doi: 10.1377/hlthaff.22.2.196.

Fagan, A.A., Bumbarger, B.K., Barth, R.P., Bradshaw, C.P., Cooper, B.R., Supplee, L.H. and Walker, D.K. (2019), "Scaling up evidence-based interventions in US public systems to prevent behavioral health problems: challenges and opportunities", *Prevention Science*, Vol. 20 No. 8, pp. 1147-1168, doi: 10.1007/s11121-019-01048-8.

Fixsen, D.L., Blase, K.A., Naoom, S.F. and Wallace, F. (2009), "Core implementation components", *Research on Social Work Practice*, Vol. 19 No. 5, pp. 531-540, doi: 10.1177/1049731509335549.

Flaspohler, P.D., Meehan, C., Maras, M.A. and Keller, K.E. (2012b), "Ready, willing, and able: developing a support system to promote implementation of school-based prevention programs", *American Journal of Community Psychology*, Vol. 50 Nos 3/4, pp. 428-444, doi: 10.1007/s10464-012-9520-z.

Flaspohler, P., Lesesne, C.A., Puddy, R.W., Smith, E. and Wandersman, A. (2012a), "Advances in bridging research and practice: introduction to the second special issue on the interactive system

framework for dissemination and implementation", *American Journal of Community Psychology*, Vol. 50 Nos 3/4, pp. 271-281, doi: 10.1007/s10464-012-9545-3.

Flay, B.R. (1986), "Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs", *Preventive Medicine*, Vol. 15 No. 5, pp. 451-474, doi: 10.1016/0091-7435(86)90024-1.

Forum on Promoting Children's Cognitive, Affective, and Behavioral Health, Board on Children, Youth, and Families, Institute of Medicine, & National Research Council (2014), *Strategies for Scaling Effective Family-Focused Preventive Interventions to Promote Children's Cognitive, Affective, and Behavioral Health: Workshop Summary*, National Academies Press (US), available at: www.ncbi.nlm.nih.gov/books/NBK230080/

Gambrill, E. (2016), "Is social work evidence-based? Does saying so make it so? Ongoing challenges in integrating research, practice and policy", *Journal of Social Work Education*, Vol. 52 No. sup1, pp. S110-S125, doi: 10.1080/10437797.2016.1174642.

Geher, G. (2017), "Why 'science-based' matters", *Psychology Today*.

Glasgow, R.E. and Emmons, K.M. (2007), "How can we increase translation of research into practice? Types of evidence needed", *Annual Review of Public Health*, Vol. 28 No. 1, pp. 413-433, doi: 10.1146/annurev.publhealth.28.021406.144145.

Glasgow, R.E., Lichtenstein, E. and Marcus, A.C. (2003), "Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition", *American Journal of Public Health*, Vol. 93 No. 8, pp. 1261-1267, doi: 10.2105/AJPH.93.8.1261.

Green, L.W., Ottoson, J.M., García, C. and Hiatt, R.A. (2009), "Diffusion theory and knowledge dissemination, utilization, and integration in public health", *Annual Review of Public Health*, Vol. 30 No. 1, pp. 151-174, doi: 10.1146/annurev.publhealth.031308.100049.

Green, J., Liem, G.A.D., Martin, A.J., Colmar, S., Marsh, H.W. and McInerney, D. (2012), "Academic motivation, self-concept, engagement, and performance in high school: key processes from a longitudinal perspective", *Journal of Adolescence*, Vol. 35 No. 5, pp. 1111-1122, doi: 10.1016/j.adolescence.2012.02.016.

Habicht, J.P., Victora, C.G. and Vaughan, J.P. (1999), "Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact", *International Journal of Epidemiology*, Vol. 28 No. 1, pp. 10-18, doi: 10.1093/ije/28.1.10.

Hall, J.N., Ahn, J. and Greene, J.C. (2012), "Values engagement in evaluation: ideas, illustrations, and implications", *American Journal of Evaluation*, Vol. 33 No. 2, pp. 195-207, doi: 10.1177/1098214011422592.

Hallfors, D. and Godette, D. (2002), "Will the 'principles of effectiveness' improve prevention practice? Early findings from a diffusion study", *Health Education Research*, Vol. 17 No. 4, pp. 461-470, doi: 10.1093/her/17.4.461.

Hecht, M.L., Graham, J.W. and Elek, E. (2006), "The drug resistance strategies intervention: program effects on substance use", *Health Communication*, Vol. 20 No. 3, pp. 267-276, doi: 10.1207/s15327027hc2003_6.

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A. (2010), "The rate of return to the HighScope perry preschool program", *Journal of Public Economics*, Vol. 94 Nos 1/2, pp. 114-128, doi: 10.1016/j.jpubeco.2009.11.001.

Jagosh, J., Bush, P.L., Salsberg, J., Macaulay, A.C., Greenhalgh, T., Wong, G., Cargo, M., Green, L.W., Herbert, C.P. and Pluye, P. (2015), "A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects", *BMC Public Health*, Vol. 15 No. 1, p. 725, doi: 10.1186/s12889-015-1949-1.

Jones, L. (2014), *Improving Efforts to Prevent Children's Exposure to Violence – a Handbook to Support Child Maltreatment Prevention Programs*, World Health Organization, Geneva.

Kelly, B. (2012), "Implementation science for psychology in education", in Kelly, B. and Perkins, D.F. (Eds), *Handbook of Implementation Science for Psychology in Education*, Cambridge University Press, Cambridge, pp. 3-12, doi: 10.1017/CBO9781139013949.003

Lachman, J., Wamoyi, J., Spreckelsen, T., Wight, D., Maganga, J. and Gardner, F. (2020), "Combining parenting and economic strengthening programmes to reduce violence against children: a cluster randomised controlled trial with predominantly male caregivers in rural Tanzania", *BMJ Global Health*, Vol. 5 No. 7, p. e002349, doi: 10.1136/bmjgh-2020-002349.

Leijten, P., Gardner, F., Melendez-Torres, G.J., Knerr, W. and Overbeek, G. (2018), "Parenting behaviors that shape child compliance: a multilevel meta-analysis", *Plos One*, Vol. 13 No. 10, p. e0204929, doi: 10.1371/journal.pone.0204929.

Lilienfeld, S.O. (2007), "Psychological treatments that cause harm", *Perspectives on Psychological Science*, Vol. 2 No. 1, pp. 53-70, doi: 10.1111/j.1745-6916.2007.00029.x.

Lyles, C.M., Crepaz, N., Herbst, J.H. and Kay, L.S. (2006), "Evidence–based HIV behavioral prevention from the perspective of the CDC's HIV/AIDS prevention research synthesis team", *AIDS Education and Prevention*, Vol. 18, pp. 21-31, doi: 10.1521/aeap.2006.18.supp.21.

Madrid, B.J., Lopez, G.D., Dans, L.F., Fry, D.A., Duka-Pante, F.G.H. and Muyot, A.T. (2020), "Safe schools for teens: preventing sexual abuse of urban poor teens, proof-of-concept study – improving teachers' and students' knowledge, skills and attitudes", *Heliyon*, Vol. 6 No. 6, p. e04080, doi: 10.1016/j.heliyon.2020.e04080.

Marchal, B., van Belle, S., van Olmen, J., Hoerée, T. and Kegels, G. (2012), "Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research", *Evaluation*, Vol. 18 No. 2, pp. 192-212, doi: 10.1177/1356389012442444.

Marsiglia, F.F., Kulis, S., Yabiku, S.T., Nieri, T.A. and Coleman, E. (2011), "When to intervene: elementary school, middle school or both? Effects of keepin' it REAL on substance use trajectories of Mexican heritage youth", *Prevention Science*, Vol. 12 No. 1, pp. 48-62, doi: 10.1007/s11121-010-0189-y.

Melendez-Torres, G.J., Leijten, P. and Gardner, F. (2019), "What are the optimal combinations of parenting intervention components to reduce physical child abuse recurrence? Reanalysis of a systematic review using qualitative comparative analysis", *Child Abuse Review*, Vol. 28 No. 3, pp. 181-197, doi: 10.1002/car.2561.

Menting, A.T.A., Orobio de Castro, B. and Matthys, W. (2013), "Effectiveness of the incredible years parent training to modify disruptive and prosocial child behavior: a meta-analytic review", *Clinical Psychology Review*, Vol. 33 No. 8, pp. 901-913, doi: 10.1016/j.cpr.2013.07.006.

Mercy, J.A., Rosenberg, M.L., Powell, K.E., Broome, C.V. and Roper, W.L. (1993), "Public health policy for preventing violence", *Health Affairs*, Vol. 12 No. 4, pp. 7-29, doi: 10.1377/hlthaff.12.4.7.

Meyers, D.C., Durlak, J.A. and Wandersman, A. (2012), "The quality implementation framework: a synthesis of critical steps in the implementation process", *American Journal of Community Psychology*, Vol. 50 Nos 3/4, pp. 462-480, doi: 10.1007/s10464-012-9522-x.

Mihalic, S.F. and Elliott, D.S. (2015), "Evidence-based programs registry: blueprints for healthy youth development", *Evaluation and Program Planning*, Vol. 48, pp. 124-131, doi: 10.1016/j.evalprogplan.2014.08.004.

Miller, R.L. and Shinn, M. (2005), "Learning from communities: overcoming difficulties in dissemination of prevention and promotion efforts", *American Journal of Community Psychology*, Vol. 35 Nos 3/4, pp. 169-183, doi: 10.1007/s10464-005-3395-1.

Milstein, R.L. and Wetterhall, S. (1999), *Framework for Program Evaluations in Public Health*, Center for Disease Control, Atlanta, GA.

Moir, T. (2018), "Why is implementation science important for intervention design and evaluation within educational settings?", *Frontiers in Education*, Vol. 3, p. 61, doi: 10.3389/feduc.2018.00061.

Morrissey, E., Wandersman, A., Seybolt, D., Nation, M., Crusto, C. and Davino, K. (1997), "Toward a framework for bridging the gap between science and practice in prevention: a focus on evaluator and practitioner perspectives", *Evaluation and Program Planning*, Vol. 20 No. 3, pp. 367-377, doi: 10.1016/S0149-7189(97)00016-5.

Mrazek, P.B. and Haggerty, R.J. (1994), *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*, National Academies Press, Washington DC.

Mullen, E.J. and Streiner, D.L. (2004), "The evidence for and against evidence-based practice", *Brief Treatment and Crisis Intervention*, Vol. 4 No. 2, pp. 111-121, doi: 10.1093/brief-treatment/mhh009.

Nathan, P.E. and Gorman, J.M. (2015), *A Guide to Treatments That Work*, Oxford University Press, Oxford.

Neuhoff, A., Loomis, E. and Ahmed, F. (2022), "What's standing in the way of the spread of evidence-based programs?", Bridgespan Group, Boston, MA.

Olds, D.L. (2006), "The nurse–family partnership: an evidence-based preventive intervention", *Infant Mental Health Journal*, Vol. 27 No. 1, pp. 5-25, doi: 10.1002/imhj.20077.

Olweus, D. and Limber, S.P. (2010), "Bullying in school: evaluation and dissemination of the Olweus bullying prevention program", *American Journal of Orthopsychiatry*, Vol. 80 No. 1, p. 124, doi: 10.1111/j.1939-0025.2010.01015.x.

Patton, M.Q. (1994), "Developmental evaluation", *Evaluation Practice*, Vol. 15 No. 3, pp. 311-319, doi: 10.1177/109821409401500312.

Patton, M.Q., McKegg, K. and Wehipeihana, N. (Eds) (2016), *Developmental Evaluation Exemplars: Principles in Practice*, The Guilford Press, New York, NY.

Peters, D.H., Adam, T., Alonge, O., Agyepong, I.A. and Tran, N. (2013), "Implementation research: what it is and how to do it", *BMJ*, Vol. 347, p. f6753, doi: 10.1136/bmj.f6753.

Richardson, T. (2009), "Challenges for the scientist-practitioner model in contemporary clinical psychology", *Psych-Talk*, Vol. 62, pp. 20-26.

Ringwalt, C.L., Ennett, S., Vincus, A., Thorne, J., Rohrbach, L.A. and Simons-Rudolph, A. (2002), "The prevalence of effective substance use prevention curricula in US middle schools", *Prevention Science*, Vol. 3 No. 4, pp. 257-265, doi: 10.1023/A:1020872424136.

Rosenbaum, D.P. and Hanson, G.S. (1998), "Assessing the effects of school-based drug education: a six-year multilevel analysis of project D.A.R.E", *Journal of Research in Crime and Delinquency*, Vol. 35 No. 4, pp. 381-412, doi: 10.1177/0022427898035004002.

Rossi, P.H., Lipsey, M.W. and Henry, G.T. (2019), *Evaluation: A Systematic Approach*, 8th ed., Sage.

Sanson-Fisher, R.W., Bonevski, B., Green, L.W. and D'Este, C. (2007), "Limitations of the randomized controlled trial in evaluating population-based health interventions", *American Journal of Preventive Medicine*, Vol. 33 No. 2, pp. 155-161, doi: 10.1016/j.amepre.2007.04.007.

Soriano, F.I. (2013), *Conducting Needs Assessments: A Multidisciplinary Approach*, Sage Publication, Thousand Oaks, CA.

Tabak, R.G., Khoong, E.C., Chambers, D.A. and Brownson, R.C. (2012), "Bridging research and practice", *American Journal of Preventive Medicine*, Vol. 43 No. 3, pp. 337-350, doi: 10.1016/j.amepre.2012.05.024.

Truman, B.I., Smith-Akin, C.K., Hinman, A.R., Gebbie, K.M., Brownson, R., Novick, L.F., Lawrence, R.S., Pappaioanou, M., Fielding, J., Evans, C.A., Guerra, F.A., Vogel-Taylor, M., Mahan, C.S., Fullilove, M. and Zaza, S. (2000), "Developing the guide to community preventive services – overview and rationale11The names and affiliations of the task force members are listed on page v of this supplement and at www.thecommunityguide.org", *American Journal of Preventive Medicine*, Vol. 18 No. 1, pp. 18-26, doi: 10.1016/S0749-3797(99)00124-5.

Valters, C. (2015), *Theories of Change. Time for a Radical Approach to Learning in Development*, Overseas Development Institute, London.

Valters, C., Cummings, C. and Nixon, H. (2016), *Putting Learning at the Centre: Adaptive Development Programming in Practice*, Overseas Development Institute, London.

Wandersman, A. (2003), "Community science: bridging the gap between science and practice with community-centered models", *American Journal of Community Psychology*, Vol. 31 Nos 3/4, pp. 227-242, doi: 10.1023/A:1023954503247.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R. and Saul, J. (2008), "Bridging the gap between prevention research and practice: the interactive systems framework for dissemination and implementation", *American Journal of Community Psychology*, Vol. 41 Nos 3/4, pp. 171-181, doi: 10.1007/s10464-008-9174-z.

Wandersman, A. and Florin, P. (2003), "Community interventions and effective prevention", *American Psychologist*, Vol. 58 Nos 6/7, pp. 441-448, doi: 10.1037/0003-066X.58.6-7.441.

Ward, C., Mikton, C., Cluver, L., Cooper, P., Gardner, F., Hutchings, J., McLaren Lachman, J., Murray, L., Tomlinson, M. and Wessels, I. (2014), "Parenting for lifelong health: from South Africa to other low- and middle-income countries", *Early Childhood Matters*, Vol. 122, available at: https://cafo.org/wp-content/uploads/2015/08/Parenting-for-Lifelong-Health1.pdf

Wathen, C.N. and MacMillan, H.L. (2018), "The role of integrated knowledge translation in intervention research", *Prevention Science*, Vol. 19 No. 3, pp. 319-327, doi: 10.1007/s11121-015-0564-9.

Weiss, C.H. (2011), "Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families", *Undefined*, available at: www. semanticscholar.org/paper/Nothing-as-Practical-as-Good-Theory-%3A-Exploring-for-Weiss/ed98a1ac4 b7b54ef4854b7b7a802db7b3e46ae02

West, S.G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D.C., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M. and Mullen, P.D. (2008), "Alternatives to the randomized controlled trial", *American Journal of Public Health*, Vol. 98 No. 8, pp. 1359-1366, doi: 10.2105/AJPH.2007.124446.

West, S.L. and O'Neal, K.K. (2004), "Project D.A.R.E. outcome effectiveness revisited", *American Journal of Public Health*, Vol. 94 No. 6, pp. 1027-1029.

Wholey, J.S., Hatry, H.P. and Newcomer, K.E. (Eds) (2010), *Handbook of Practical Program Evaluation*, 3rd ed., Wiley & Sons, New York, NY.

Wiseman, S.H., Chinman, M., Ebener, P.A., Hunter, S.B., Imm, P. and Wandersman, A. (2007), "Getting To Outcomes[TM] [product page]", available at: www.rand.org/pubs/technical_reports/TR101z2.html

Zaza, S. Briss, P.A. and Harris, K.W. (2005), "The guide to community preventive services: what works to promote health?", available at: www.cabdirect.org/cabdirect/abstract/20053043169

Corresponding author

Patricia Lannen can be contacted at: lannen@mmi.ch