ADVANCING EVALUATION
OF CHARACTER BUILDING PROGRAMS

Jennifer Brown Urban
Montclair State University

William M. Trochim
Cornell University

This article presents how character development practitioners, researchers, and funders might think about evaluation, how evaluation fits into their work, and what needs to happen in order to sustain evaluative practices. A broader view of evaluation is presented whereby evaluation is not just seen as something that is applied at a program level, but as an endeavor that considers the ecologies and systems within which programs are embedded. The evaluation landscape for youth-serving and character building programs is considered. Strategies for enhancing evaluation practices at the organizational (macro) level are addressed as well as strategies for enhancing evaluation practice at the program (micro) level. Organizational level strategies include deliberate efforts to engage in evaluation capacity building and cultivating evaluative thinking. We also discuss the role of evaluation policy and the need to consider program ecosystems through portfolio analysis. The program level strategies focus on using the Systems Evaluation Protocol to articulate and assess program theories of change.

Perhaps it is best to start out by describing what this article is not going to do. It will not provide an overview of evaluation, describe major approaches, delineate the broad range of available methodologies, or provide examples of evaluations relevant to the topic of character development programs. Put more succinctly, we do not intend to characterize how evaluation is done. The mainstream literature on evaluation addresses all of these topics far better than we could hope to do here. We assume that either the reader is familiar with that literature or, if interest so motivates, will be able to avail themselves of it at any point. As such, the intended audience for this article are readers who have at least some basic knowledge of evaluation.

Instead this article hopes to make a more general argument about how character development practitioners, researchers, and funders might think about evaluation, how it fits into their work, and what needs to happen in order to sustain it. Along the way we hope to introduce the reader to a broader view of evaluation, not just as something that is applied to a program or at a program level but as an endeavor

---

• **Correspondence concerning this article should be addressed to:** Jennifer Brown Urban, urbanj@montclair.edu

---

that needs to be integral to the ecologies and systems within which programs occur. This macrolevel view is driven by several tectonic shifts in high-level thinking in evaluation (and many other fields) over the past few decades associated with the rise of ecological and systems thinking, the integration of evolutionary theory into applied social research methodology, and the rise of integrated global computer technology in the form of the Internet. These forces are moving us away from the traditional focus of evaluation at the program level and situating traditional program evaluation within a longer term evolution of program theory and practice. It is at this macroevolutionary level, rather than at the within-program level, that we wish to focus this article and the field of character development.

We begin at the most familiar level, briefly describing the landscape of the evaluation of youth-serving and character building programs. We then discuss strategies for enhancing evaluation practices at both the organizational (macro) level followed by a discussion of strategies for enhancing evaluation practices at the program (micro) level.

Our hope for this article is that it will move the discussion about evaluation, if even slightly, to a higher and more systemic level in the field of character development, move us away from the focus on the evaluation of programs and toward a more evolutionary evaluation perspective focused on evidence-driven program theory development and the assessment of theoretically described portfolios of ideas about positive youth development toward what we hope will be a deeper understanding of character development and how it might be influenced and encouraged.

HISTORY OF EVALUATION IN POSITIVE YOUTH DEVELOPMENT AND CHARACTER BUILDING PROGRAMS

When we survey the landscape of youth-serving programs, including both in-school and out-of-school programs, we can readily see that many are designed to promote character attributes. However, the vast majority of these programs have not been formally evaluated and we have little idea of whether and how they work (Roth & Brooks-Gunn, 2003, 2016; Roth, Brooks-Gunn, Murray, & Foster, 1998). Character development for youth is increasingly being recognized as critically important; however, our ability to evaluate and bring effective programs to scale has not kept pace. Early work in youth/character development emphasized shifting the focus from a deficit to a strengths-based perspective. At the turn of the 21st century, the focus shifted again to advancing theory in youth/character development and providing preliminary evidence that youth-serving programs are indeed beneficial (Roth & Brooks-Gunn, 2016). We are now ready to embark on the next phase of research and evaluation of youth-serving programs, and this must include detailed and specific articulation of program theory, more careful attention to definitions and measurement, and most importantly, investment of time and resources in the planning, implementation and utilization of evaluations of such programs.

It is important to recognize that many youth-serving programs are small, local programs (e.g., sports teams, school newspaper) that are locally sponsored, while other youth-serving programs are affiliates of national organizations such as Boy/Girl Scouts or 4-H (Roth & Brooks-Gunn, 2003). Given a relatively recent focus on evaluation and the high turnover of program staff, many of these programs lack the time and financial resources as well as the skillset to engage in high quality evaluation and planning. As a result, many youth-serving programs, rely on inappropriate data or lack sufficient data to demonstrate the programs' effectiveness at promoting character. In order to increase the quality of character development programs and, in turn, increase the positive impact of these programs on youth, the field of character development must: (1) advance at the organizational level by fully integrating evaluation throughout the organizational culture; and, (2) advance evalu-

ation practice at the programmatic level by adopting an Evolutionary Evaluation perspective.

ADVANCING CHARACTER DEVELOPMENT EVALUATION

Evolutionary Evaluation considers the complex larger systems within which programs are embedded (Trochim et al., 2012; Urban, Hargraves, Hebbard, Burgermaster, & Trochim, 2011; Urban, Hargraves, & Trochim, 2014; Urban & Trochim, 2009). The Systems Evaluation Protocol (SEP) is an evaluation approach that applies principles of Evolutionary Evaluation in order to provide a foundation for planning and conducting evaluations, developing and improving programs, and building programmatic and organizational evaluation capacity (Buckley, Archibald, Hargraves, & Trochim, 2015; Trochim et al., 2012; Urban et al., 2011; Urban et al., 2014; Urban & Trochim, 2009). Evolutionary Evaluation and the SEP integrate principles from developmental systems theory (e.g., Lerner, 2006; Overton, 2006, 2010), systems theory (Bertalanffy, 1995; Laszlo, 1996; Midgley, 2003; Ragsdell, West, & Wilby, 2002), evolutionary theory (Darwin, 1859; Mayr, 2001), and evolutionary epistemology (Bradie & Harms, 2006; Campbell, 1974, 1988; Cziko & Campbell, 1990; Popper, 1973, 1985).

Evolutionary Evaluation and the SEP have important practical implications for real-world application. In this section, we turn our attention to the implications of Evolutionary Evaluation and the SEP for the evaluation of character development programs. We first consider the "global" level, the macro level of the organization and its context, including its multiple niches and cultures, its varied value systems, and how these relate to evaluation. We argue that issues of motivation, the challenges of sustaining evaluation, the need to cultivate an evaluation culture and the integral role of critical evaluative thinking are essential at this level. We consider several systems

methodologies that are essential to encouraging evolution of knowledge of character development program theory and practice. We then turn to the "local" level, the microlevel of the program and its planning, implementation and evaluation. We present the SEP which has been developed and tested over the past decade and is designed to help any program stakeholders develop a program evaluation that is systems and context sensitive, can be used to evaluate a program at any stage of its lifecourse (ontogeny) and will contribute to a more general agenda of knowledge evaluation (phylogeny). We also introduce an evolving web-based cyberinfrastructure that is freely available and can be used to manage both local program evaluations and more global portfolios of programs.

ORGANIZATIONAL (MACRO) LEVEL STRATEGIES

The challenges of building and sustaining evaluation at an organizational level are many, particularly within the context of character development programs. In order to truly advance, the field of character development needs to commit to an enhanced focus on evaluation as a valuable, institutionalized, and embedded aspect of organizational practice (Mayne, 2010; Sanders, 2002). This includes adequately investing time and money in evaluation and evaluation planning. Evaluation should not be an afterthought, but rather should be considered early on in program planning and ideally should be fully embedded within program development and practice. Equally important is the need for committed leadership, especially at the funder level in order to build and sustain evaluation systems. Given the considerable turnover in frontline staff for character development programs, the need for sustainable evaluation systems and policies is even more critical. Specifically, the field of character development needs to work toward building an evaluation culture.

Systems Change

This section describes the areas where broad-level systems change typically needs to happen in order to develop a sustainable and effective evaluative function in a complex hierarchical system like the character development context. We provide brief discussions of evaluation capacity building and evaluative thinking and encourage the interested reader to refer to our research groups' more in-depth discussions of these topics in Archibald, Sharrock, Buckley, and Cook (2016) and Buckley et al. (2015).

Evaluation Capacity Building. In the past 15 years, the field of evaluation has seen a proliferation of evaluation capacity building definitions, models, and approaches. The most commonly cited definition of evaluation capacity building is "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine" (Stockdill, Baizerman, & Compton, 2002, p. 14). Evaluation capacity building is often associated with collaborative, participatory, and empowerment evaluation (Fetterman & Wandersman, 2005; O'Sullivan, 2004; Rodriguez-Campos, 2005) all of which share a common interest in democratizing and decentralizing evaluation practice. This stands in contrast to more traditional models of program evaluation wherein evaluation was solely the purview of expert researchers. Evaluation capacity building provides opportunities to infuse organizations and programs with evaluation skills, attitudes, and practices that promote a culture of evaluation and ultimately improve program outcomes (Labin, 2014; Suarez-Balcazar & Taylor-Ritzler, 2014; Wandersman, 2014).

Systems that consist of multiple projects and organizations typically require a wide range of evaluation skills, resources and capabilities in order to provide the essential feedback processes for system learning and to meet the accountability reporting needs of funders and stakeholders. Some of these needs are best met by developing internal organizational or system capacity by hiring experienced evaluators, supporting external or in-service training in evaluation, hiring external consultants, and so on. Evaluation capability is sometimes met by funders who require evaluation, through additional funding for this function or through evaluation technical assistance. In many cases, individuals or groups within organizations simply seek out resources and information on evaluation, increasingly these days through technology-based solutions such as evaluation websites. However, there is little in the way of systematizing the plethora of resources and little support for practitioners who need to navigate this complex terrain.

Character development program practitioners and funders can deliberately work to enhance the internal evaluation capacity of program staff by providing the time and resources to engage in evaluation activities as a core function of program work. Some organizations have created internal evaluation roles or departments that focus on program monitoring and evaluation. Other organizations have invested in professional development for some client-serving staff to build their knowledge and skill set around evaluation. In order for character development programs to make significant progress, evaluation needs to be prioritized.

Evaluative Thinking. One of the primary ways to build evaluation capacity is by developing an organizational culture of evaluation where evaluative thinking permeates all planning, monitoring, and evaluation processes (Archibald et al., 2016). A key development in the field of evaluation is the idea of evaluative thinking as an essential skill for evaluation and as a process for building evaluation capacity (Compton, Baizerman, & Stockdill, 2002). At an organizational level, in order to build evaluation capacity, developing and promoting an organizational culture that supports evaluative thinking is critical. Several authors have defined evaluative thinking (Chinnock, 2008; Davidson, 2005; Patton, 2005; Preskill, 2008). Most agree that it includes higher order thinking and decision

making skills as well as an evaluative attitude and set of beliefs. Patton (2005) defined evaluative thinking as

> a willingness to do reality testing, to ask the question: how do we know what we think we know? To use data to inform decisions. Evaluative thinking is not just limited to evaluation projects, it's not even just limited to formal evaluation; it's an analytical way of thinking that infuses everything that goes on.

Character development program implementers who are also evaluative thinkers will be good evaluators as well as better program planners, managers, organizers and even participants. The supposition is that engaging people in evaluative thinking is also likely to be a personally exciting endeavor that enhances their understanding of and motivation to engage in evaluation, and consequently contributes to evaluation capacity and sustainability in an organization.

Character development program staff can work toward developing a culture of evaluative thinking within their organization. One strategy is to intentionally develop an evaluative thinking learning environment by, for example, highlighting lessons learned from both program successes and failures. Role playing can be used during evaluation planning to help understand various stakeholders' perspectives. Engaging in critical peer review can help highlight assumptions or leaps in logic. Perhaps most importantly, all members of the organization can be engaged in some aspect of the evaluation process (Buckley et al., 2015).

Systems Methods

Here we consider two methodological approaches that are appropriate at the macrolevel for supporting and sustaining the evaluation function in character development programs. We provide brief discussions of evaluation policy methods and funder portfolio analysis and encourage the interested reader to refer to our more in-depth discussions of these topics in Trochim (2009), Urban et al. (2014), and Urban and Trochim (2009).

Evaluation Policy Methods. Evaluation policy does not just happen on its own; it has to be developed. Trochim (2009) offered a taxonomy of eight evaluation policy domains or categories organized into a visual "policy wheel" for analyzing and managing evaluation policies (Figure 1) that includes policy domains that cover: evaluation goals; participation; capacity building; management; roles; process and methods; use; and, meta-evaluation. The policy wheel organizes evaluation policy into a simple circle diagram, divided into wedges that correspond to the proposed eight types of evaluation policies in the taxonomy. All evaluation policies can be placed somewhere on the wheel. There are different concentric circles on the wheel, with more general policies in each category placed on the outer circles and more specific policies (and ultimately practices or procedures) more central. There are also different layers in the wheel representing different levels of organizational hierarchy. The different levels correspond to different global-local arrangements in a nested hierarchy. For instance, Level 1 might correspond to a national funder level, Level 2 to a cross-site lead organization level and Level 3 to an individual program site level. Each level inherits policy from above and delegates responsibility for greater policy specificity to those below.

Trochim (2009) also describes a number of principles or rules that guide the organization of evaluation and functioning of policies in this structure and constitute the basis of a methodology for evaluation policy analysis:

- all policies and practices "inherit" their parent characteristics (inheritance);
- a child policy can never be broader than its parent (encapsulation);
- Inner policies are more specific than outer (parent) policies (specificity);
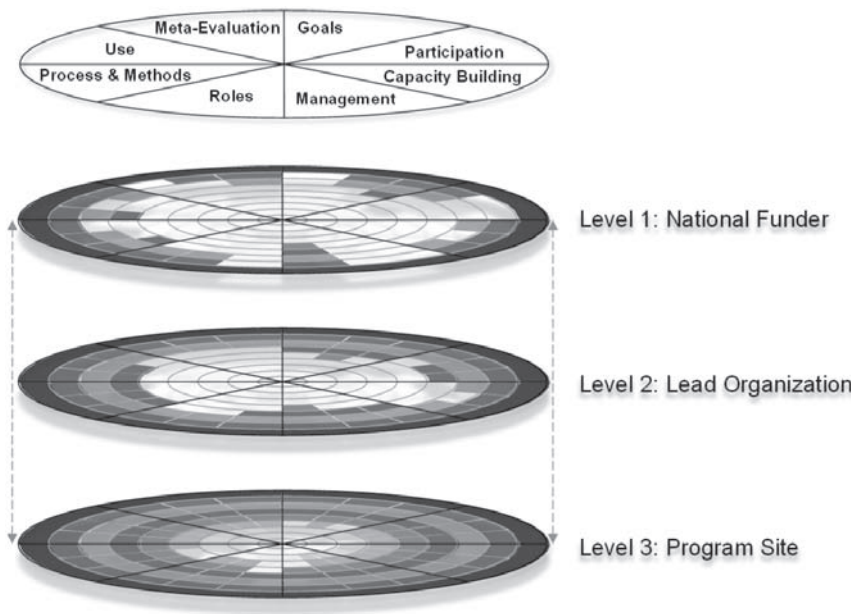- policies should cover the entire relevant domain (exhaustiveness);

FIGURE 1
The Evaluation Policy Wheel

- there should be no large gaps between levels of policies (continuity) (discontinuity suggests micromanagement);
- responsibility is delegated for more detailed policy or practice than specified (delegation); and
- "any reasonable interpretation" of the delegated policies is legitimate (accountability).

Evaluation policymaking in this model is a very dynamic and iterative process. It begins with the evaluation policymaker(s) describing the most general and highest level policies in each of the eight domains. For example, for the evaluation capacity domain, the broadest level policy might be something like: *The organization will develop and implement sufficient organizationwide capacity to support evaluation activities.* If this were the only policy in this domain it means that the policymakers are delegating to those responsible for enacting the policy the responsibility for defining and achieving it. The practitioners or staff are responsible for reporting on how they operationalized and addressed the policy. The policymakers are responsible for reviewing and evaluating the degree to which the practices that were enacted addressed their intent. Among the appealing features of this evaluation policy model are that it is inherently delegative, discourages micromanagement, and acknowledges that there may be multiple ways to address any policy and tailor it to the varying needs of organizational or program components. Organizations engaged in character development programming should consider the potential utility of using a policy model like this one for developing, managing and disseminating evaluation policy.

Funder Portfolio Analysis. Throughout most of its history, educational evaluation has tended to focus primarily on the program as the primary unit, almost as though programs exist in a systems vacuum. The literature is replete with discussions that address program evaluation without attending to the systems issues within which programs are situated. Often it is

the system that determines how evaluations are done, who is allowed to do them, how they will be resourced, when and to whom results will be reported, and so on. Yet evaluators tend to treat each evaluation as though it can be designed as a unique one-off endeavor based only or primarily on local considerations, proximal needs and immediate concerns about threats to validity.

A challenge for character development programs is that local character development programs are typically most concerned with the experiences of their participants and how service delivery can be improved (local level concerns). Funders are typically more concerned with how their portfolio of programs affect more global questions related to longer term impact (global level concerns). We need to recognize that this is essentially a hierarchical systems challenge—how to harmonize the local model of a specific program with the hierarchically broader model of programs of that type.

When evaluations do occur, they are typically done in isolation at the local level which results in a portfolio of similarly oriented programs that exist across the same or multiple organizations that often have separate mutually uninformed program models. In addition, there is also usually at least an implicit model at the next level up in the organization. For instance, in a case where a foundation funds a portfolio of grants that each have character development programs, it is likely that there is some model at the foundation level, although this may not have been formally articulated. In this case, there is typically no linkage or integration of the models either horizontally (local–local) or vertically (local–global). Programs do not identify common activities or how their activities connect with a global model. They do not harmonize their outcomes or even a subset of them, making it difficult if not impossible to aggregate results subsequently across the portfolio of programs. There is seldom formal policy across the portfolio regarding how the programs should be evaluated. One option is to impose a single

common model onto the local programs, but this runs the risk of increasing local burden, reducing flexibility and local adaptation, not being responsive to local conditions, and creating program "monocultures" (Trochim, 2007).

Instead, we propose a process whereby the local and global entities work as a system to coordinate and harmonize what they are doing, understand each other's needs and perspectives, seek common connections in their models, and encourage an emergent harmonization that both enables local flexibility and some global synthesis of results. This linked and integrated system model of programs, created together by multiple levels of the system, would provide the foundation for development and analysis of program portfolios. This becomes especially important when funders are trying to understand where they are making programmatic investments, whether their portfolio of funded work is addressing the overarching mission of the agency, where gaps may exist, and in identifying emerging areas that warrant additional investment. Without a methodology for portfolio analysis, funders typically must rely on the disjointed evaluation efforts of their grantees and post hoc evaluations that try to assess broad programmatic impact based on limited data. A fundamental issue driving the difficult process of evaluating portfolios is that at the funder level, models are typically not articulated. Without a well-articulated funder-level model, practitioners of programs at the local–level may believe they are addressing the goals and objectives of the funder, but they are unlikely to be able to describe *explicitly* how they are doing so.

An Evolutionary Evaluation approach addresses this need by building hierarchical nested program models (nested pathway models) that integrate the logic of individual programs with other peer programs and the logic of higher levels of the system that fund and oversee the programs. By developing nested program models, funders can then more effectively assess a "portfolio" or hierarchical aggregation of similar programs across a system. See the discussion on the creation of path-

way models in the section below on program (micro) level approaches.

PROGRAM (MICRO) LEVEL STRATEGIES

In this section we describe how the character development field can enhance the evolution of our knowledge about programs and what works. At the program level, we consider two ways an Evolutionary Evaluation approach can enhance our work: through the use of the SEP which integrates evolutionary and systems principles into program evaluations; and through the development of methods that integrate practice and research.

The SEP

The SEP emphasizes: the importance of creating a causal diagram that illustrates the programmatic theory of change; incorporating the perspectives of both internal and external stakeholders of the program; recognizing how the program is related to other programs either in the same system or other systems, in part by identifying research on similar or related outcomes which can help link the program to more universal long-term goals; and continually assessing and revising the theory of change and evaluation plans based on knowledge gained through evaluation efforts (Urban et al., 2014; Urban & Trochim, 2009). The SEP integrates the three primary phases of any evaluation: planning, implementation, and utilization. Each stage of each phase has a detailed list of components or steps that can be followed to accomplish an evaluation. This article emphasizes the planning phase of evaluation (Figure 2).

The SEP is designed to generate evaluation plans. It is a series of repeatable steps that when followed lead to the creation of program logic and pathway models and an evaluation plan that can subsequently be implemented and utilized. In this sense it addresses the needs of character development programs for
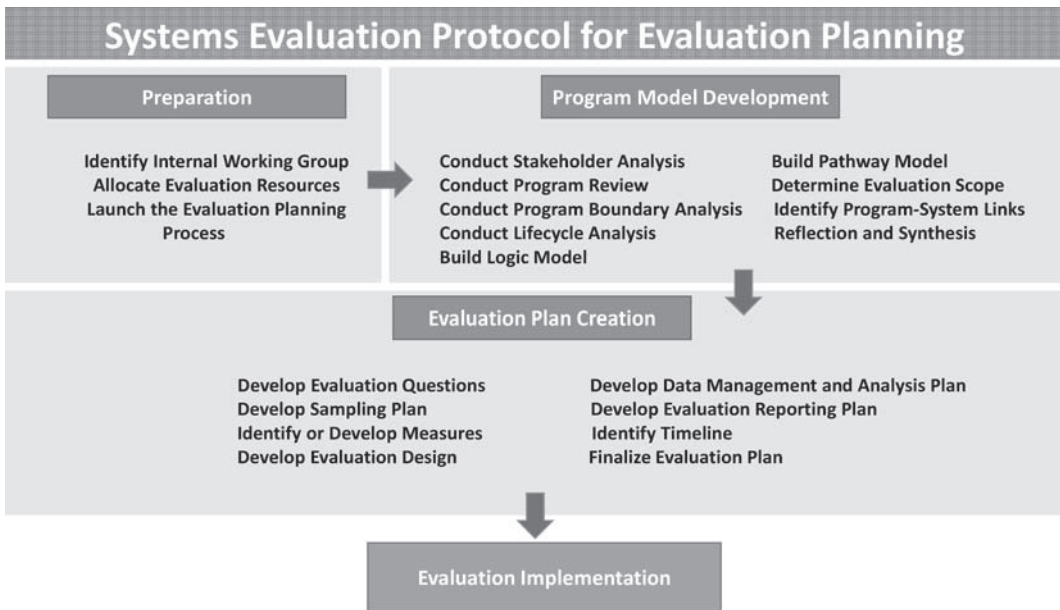

FIGURE 2
Steps of the SEP for Evaluation Planning

standardization of evaluation approaches while recognizing the enormous varieties of contexts within which character development programming occurs.

While the SEP can be implemented as a manual process that does not depend on any specific technology platform, it is designed so that it can be enhanced throughout by using a system developed in our research called the Netway, a web-based application consistent with second-generation web-based communities and hosted services such as social-networking sites, wikis, and blogs (Wikipedia, 2007). The Netway is constructed so that when program practitioners or evaluators enter program information about activities, outputs and outcomes, the system can immediately identify and suggest other existing programs that have similar or common elements and enable the users to adopt or adapt these elements for their own programs while automatically creating networked linkages of their models with others (Asim, Essegaier, & Kohli, 2000; Burke, 2000). Each new program model adds to the online network of such models and can be accessed in turn by others. This also helps ensure that different parts of the system can learn from each other and that even programs with no direct contact with one another, can use the cyberinfrastructure to benefit from each other's experiences (Marathe, 1999). Evaluators who are supporting programs that use the Netway can see their portfolio of programs and what they are adding to the system in real-time, and can communicate with program practitioners about their models and evaluation plans, thus enabling new models of virtual consultation. The system is designed so that researchers will be able to identify clusters of programs that are in their substantive areas of interests and learn about new and emerging programs that are responding to local needs and conditions. Funders (e.g., National Institutes of Health) can view meta-summaries of programs across program areas, see where they are in their developmental life cycles, and more effectively manage their portfolios of evaluations. The Netway cyberinfrastructure is

a creative incorporation of technology that fundamentally changes the nature of evaluation practice for both the evaluator and the practitioner and has the potential to be a transformative process for character development evaluation particularly and for evaluation generally in the 21st century.

SEP STAGES

The planning stage of the SEP includes three phases: (1) preparation, (2) model development, and (3) evaluation plan development. The primary objectives of the preparation stage are to acquaint the working group with the SEP process, identify people's key roles, and collect basic information about the program. The model development stage is a central and distinguishing component of the SEP, focused on surfacing and articulating deeper understandings of the program through: stakeholder analysis and mapping (a visual depiction of the stakeholders and their relationship to each other); group discussion and program review including development of a written program description; identification of program and evaluation life cycle phases; structured program modeling in two forms—the more familiar columnar logic model and a corresponding pathway model; linking the pathway model with the evidence base; and, determining the scope of the evaluation. A stakeholder map, life cycle determinations, and the logic and pathway models are the products of the model development stage. These products form the foundation for strategic decision-making about the evaluation scope, evaluation purpose, specific evaluation questions and other components of the evaluation plan which in turn serve as the basis for evaluation plan development in the third and final stage of the SEP. In the evaluation plan development stage the specific design, sample, measures, and analysis plan are developed with careful consideration of program and evaluation life cycle alignment and principles of Evolutionary Evaluation. A clear benefit of working through

the SEP is that the process: builds evaluation capacity in terms of specific skills and knowledge (modeling, life cycle analysis, evaluation methodology, etc.); deepens staff understanding of the program and ability to communicate with stakeholders; and, cultivates skills, patterns of thought, and commitment to evaluation that constitute evaluative thinking (Urban et al., 2015.

Pathway Models for Integrating Research and Practice. Local character development practitioners tend to be most interested in shorter term outcomes and improving their practice. The practitioners' dilemma is that they operate on a local level yet they are asked to demonstrate effects on long-term, broader outcomes. Those who are situated at a more global level of the system (e.g., character development program funder level) tend to be more interested in longer term outcomes and focus on broad impact. The program director's dilemma is that they are beholden to their funders and expected to demonstrate large-scale impact often aggregated across multiple local program sites. This is a classic systems thinking problem, a part–whole or local–global challenge (Young, 1999). How do we connect the varied local experiences with the broader global outcomes of interest? At the core of this systems challenge is the central role that program evaluation, planning, the research evidence base, and particularly detailed and clearly articulated program modeling can play in making these local–global connections.

The heart of the solution to this systems challenge is pathway models which articulate program theory and 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). A well-articulated pathway model (Figure 3) provides the foundation for systematically and dynamically linking program theory with the research evidence base. 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 by graphically depicting a network of causal linkages, primary pathways and nodes.

The SEP facilitates research-practice integration particularly during the process of linking the pathway model with the research evidence base. Combining the research literature with high-quality evaluation can help build a strong case for the underlying theory of change. We provide an example from the Inspire Aspire character development program in which youth reflect on personal strengths and areas in need of improvement, research an inspirational figure, and contemplate what they can do to bring their vision for a better world to life. The program culminates with the youth creating a poster. The interested reader is referred to Urban, Linver, Thompson, Davidson, and Lorimer (2017) for an in-depth discussion and case study of the SEP's successful application with Inspire Aspire.

Figure 4 zooms in on the first highlighted throughline from the pathway model in Figure 3. In our evaluation of Inspire Aspire, we knew that we would not be able to follow the youth long enough to actually measure the long-term outcome of interest. However, we were able to measure the short- and medium-term outcomes. We developed a tool to code student posters for alignment between values and articulation of future ambitions (short-term outcome). We went to the research literature and found validated scales for measuring youth goal setting (short-term outcome; Freund & Baltes, 2002; Gestsdóttir, Bowers, von Eye, Napolitano, & Lerner, 2010) and we measured whether there was an association between better aligned posters (short-term outcome) and scores on goal setting measures (short-term outcome). We also found scales that measure sense of purpose (Bundick et al., 2008; Steger, Frazier, Oishi, & Kaler, 2006) and measured whether there was an association between goal setting (short-term outcome) and sense of purpose (medium-term outcome). However, we
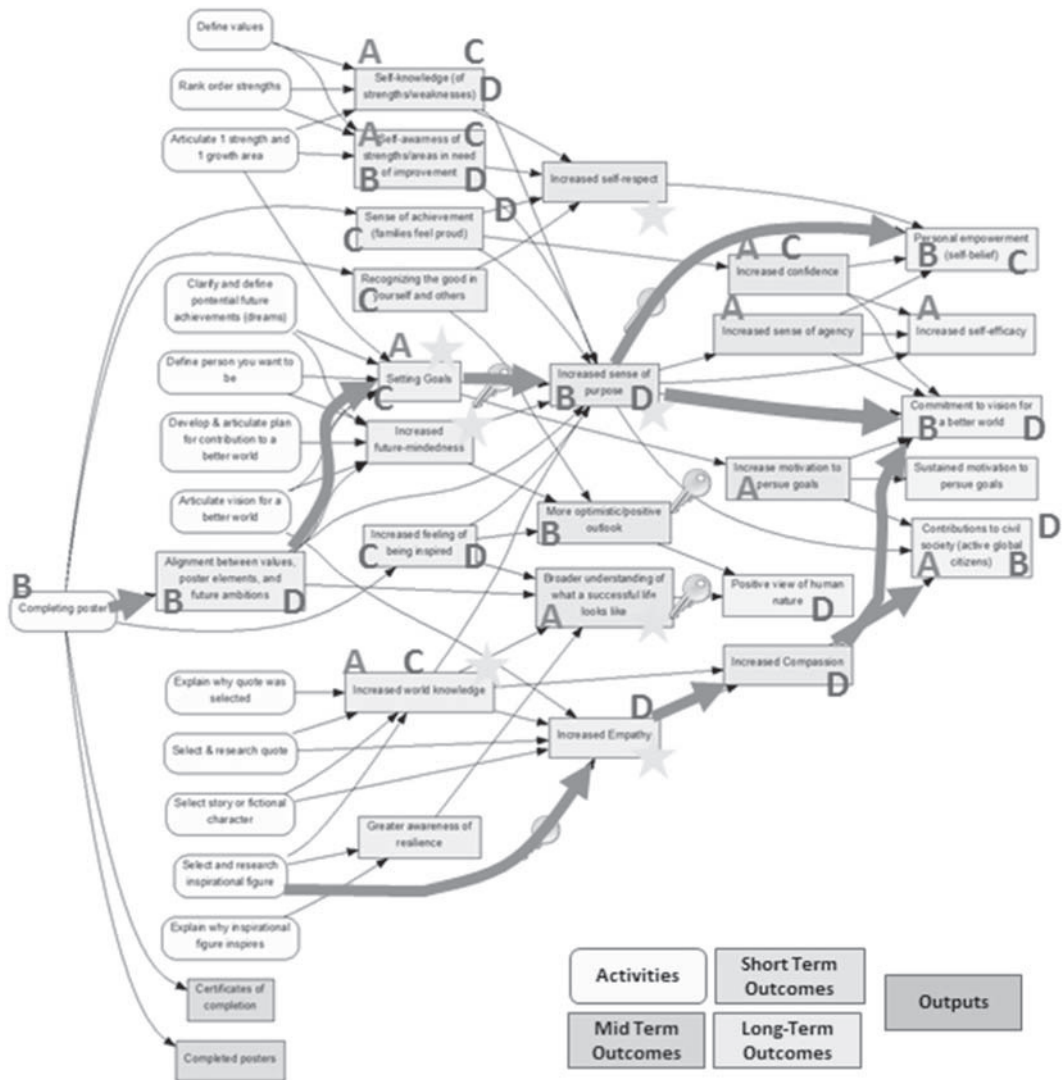
FIGURE 3
Inspire>Aspire Pathway Model After Completing Mining the Model
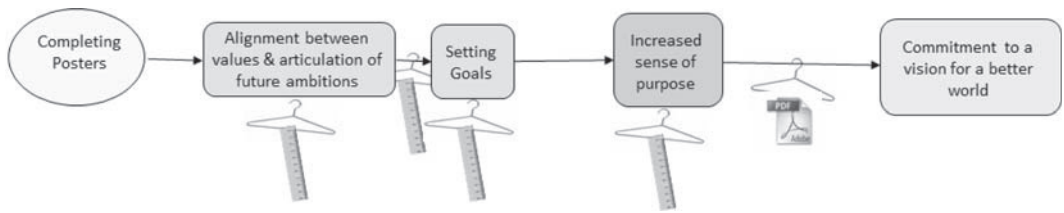

FIGURE 4
Illustrations of the Golden Spike

did not have resources to follow-up with these kids in the future to see if they also go on to develop a commitment to a vision for a better world (long-term outcome); but we have found research that shows that increased sense of purpose and goal setting abilities lead to a commitment to a vision for a better world (Moran, 2014; Yeager, Bundick, & Johnson, 2012). The research evidence-base can pick up where we left off in our measurement and demonstrate the logical connection to longer term outcomes. Ultimately, our goal is to link the evidence that is derived from a "local" evaluation of a program with the more "global" evidence that is generated by research, to identify places where local evaluation efforts and the research literature meet; in other words, to find what we refer to as the "golden spike" (Urban & Trochim, 2009). The development of the pathway model and the subsequent linkage with the research evidence base provides a framework for identifying and supporting the connections from short- and medium-term outcomes to long-term ones. A pathway model 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.

CONCLUSIONS

Evaluation is essential to the continued growth of the field of character development and Evolutionary Evaluation which is aligned with Relational Developmental Systems meta-theory is particularly well-suited to the complex challenges associated with evaluation of character development programs and systems. This systems thinking perspective takes into account the dynamic nature of character development programs as well as the dynamic processes character development programs aim to address. The field of character development is at a crossroads and it is time to make significant investments in evaluation in order to reach our shared goal of increasing the number of flourishing people committed to enhancing civil society. This must be done at both the macro, organizational level, as well as at the more micro, program level. Funders must be willing to invest in building evaluation systems. Organizations must be willing to invest time and energy in building evaluation capacity and evaluation policies. Program leaders and practitioners must be committed to careful evaluation planning. And, people at all levels of the system must be willing to understand the broader system, their place within it, and the shared commitment needed to make meaningful change. With these goals in mind, we can collectively advance the evaluation of character development programs.

NOTE

1. The Netway is currently available as a free web service at http://www.evaluationnetway.com/ that requires the user to register with a unique Username and Password. The Netway incorporates extensive resources including the complete documentation of both the SEP and the web-based system itself.

REFERENCES

Archibald, T., Sharrock, G., Buckley, J., & Cook, N. (2016). Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development. *Evaluation and Program Planning, 59*, 119–127.

Asim, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research, 37*(3), 363–376.

Bertalanffy, L. V. (1995). *General system theory: Foundations, development, applications* (Rev. ed.). New York, NY: Braziller.

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 33*, 5–18.

Bradie, M., & Harms, W. (2006). Evolutionary epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2016 ed.). Stanford, CA: The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on

critical thinking. *American Journal of Evaluation, 36*(3), 375–388.

Bundick, M., Andrews, M., Jones, A., Mariano, J. M., Bronk, K. C., & Damon, W. (2006). *Revised youth purpose survey.* Stanford, CA: Stanford Center on Adolescence.

Burke, R. (2000). Knowledge-based recommender systems. In A. Kent (Ed.), *Encyclopedia of library and information systems* (Vol. 69, Supplement 32). New York, NY: Marcel Dekker.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). LaSalle, IL: Open Court.

Campbell, D. T. (1988). Evolutionary epistemology. In E. S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers of Donald T. Campbell* (pp. 393–434). Chicago, IL: University of Chicago Press.

Chen, H., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283–302.

Chinnock, P. (2008). Knowledge translation toolkit: A resource for researchers. Retrieved July 30, 2010, from http://www.tropika.net/svc/specials/KT-Toolkit/pages/KT-Toolkit

Compton, D. W., Baizerman, M., & Stockdill, S. H. (Eds.). (2002). *The art, craft, and science of evaluation capacity building* (Vol. 93). San Francisco, CA: Jossey Bass.

Cziko, G. A., & Campbell, D. T. (1990). Comprehensive evolutionary epistemology bibliography. *Journal of Social and Biological Structures, 13*(1), 41–82.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London, England: John Murray.

Davidson, E. J. (2005). *Evaluative thinking and learning-enabled organizational cultures.* Paper presented at the Canadian Evaluation Society & American Evaluation Association Conference, Toronto, Ontario.

Fetterman, D., & Wandersman, A. (Eds.). (2005). *Empowerment evaluation principles in practice.* New York, NY: Guilford Press.

Freund, A. M., & Baltes, P. B. (2002). Life-management strategies of selection, optimization and compensation: Measurement by self-report and construct validity. *Journal of Personality and Social Psychology, 82*(4), 642–662.

Gestsdóttir, S., Bowers, E., von Eye, A., Napolitano, C. M., & Lerner, R. M. (2010). Intentional self regulation in middle adolescence: The emerging role of loss-based selection in positive youth development. *Journal of Youth and Adolescence, 39*(7), 764–782.

Labin, S. N. (2014). Developing common measures in evaluation capacity building an iterative science and practice process. *American Journal of Evaluation, 35*(1), 107–115.

Laszlo, E. (1996). *The systems view of the world: A holistic vision for our time (Advances in systems theory, complexity, and the human sciences).* Creskill, NJ: Hampton Press.

Lerner, R. M. (2006). Developmental science, developmental systems, and contemporary theories of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., Vol. 1, pp. 1–17). Hoboken, NJ: John Wiley & Sons.

Marathe, J. (1999). *Creating community online.* Durlacher Research. Retrieved from http://www.delijst.net/delijst/pdf/unpan003006.pdf

Mayne, J. (2010). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation, 24*(2), 1–30.

Mayr, E. (2001). *What evolution is.* New York, NY: Basic Books.

McLaughlin, J., & Jordan, G. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning, 22*(65–72).

Midgley, G. (2003). *Systems thinking.* Thousand Oaks, CA: SAGE.

Moran, S. (2014). What "purpose" means to youth: Are there cultures of purpose? *Applied Developmental Science, 18*(3), 163–175. doi:10.1080/10888691.2014.924359

O'Sullivan, R. G. (2004). *Practicing evaluation: A collaborative approach.* Thousand Oaks, CA: SAGE.

Overton, W. F. (2006). Developmental psychology: Philosophy, concepts, methodology. In R. M. Lerner (Ed.), *Theoretical models of human development* (6th ed., Vol. 1, pp. 18–88). Hoboken, NJ: Wiley.

Overton, W. F. (2010). Life-span development: Concepts and issues. In W. F. Overton & R. M. Lerner (Eds.), *Handbook of life-span development: Cognition, biology, and methods* (Vol. 1 pp. 1–29). Hoboken, NJ: Wiley.

Patton, M. (2005). Interview with Lisa Waldick, from the International Development Research

Center. Retrieved August 25, 2005, from https://www.idrc.ca/en/article/conversation-michael-quinn-patton

Popper, K. (1973). *Evolutionary epistemology*. Paper presented at the Sections I–VI of "The Rationality of Scientific Revolutions" given at the Herbert Spencer Lecture, University of Oxford.

Popper, K. (1985). Evolutionary epistemology. In D. M. Miller (Ed.), *Popper selections* (pp. 78–86). Princeton, NJ: Princeton University Press.

Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation, 29*(2), 127–138.

Ragsdell, G., West, D., & Wilby, J. (2002). *Systems theory and practice in the knowledge age*. New York, NY: Kluwer Academic/Plenum.

Rodriguez-Campos, L. (2005). *Collaborative evaluations*. Tamarac, FL: Llumina Press.

Roth, J. L., & Brooks-Gunn, J. (2003). Youth development programs: Risk, prevention, and policy. *Journal of Adolescent Health, 32*(3), 170–182.

Roth, J. L., & Brooks-Gunn, J. (2016). Evaluating youth development programs: Progress and promise. *Applied Developmental Science, 20*(3), 188–202. doi:10.1080/10888691.2015.1113879

Roth, J. L., Brooks-Gunn, J., Murray, L., & Foster, W. (1998). Promoting healthy adolescents: Synthesis of youth development program evaluations. *Journal of Research on Adolescence, 8*(4), 423–459.

Sanders, J. R. (2002). Presidential address: On mainstreaming evaluation. *American Journal of Evaluation, 23*, 253–259.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning of life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling, 53*(1), 80–93.

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In D. W. Compton, M. Baizerman & S. H. Stockdill (Eds.), *The art, craft and science of evaluation capacity building: New directions for evaluation* (Vol. 93, pp. 7–25). San Francisco, CA: Wiley.

Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2014). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation, 35*(1), 95–99.

Trochim, W. (2007). *Evolutionary perspectives in evaluation: Theoretical and practical Implications*. Paper presented at the 30th Annual Conference of the Eastern Evaluation Research Society, Absecon, NJ. Retrieved from http://www.socialresearchmethods.net/research/EERS2007/Evolutionary%20Perspectives%20in%20Evaluation%20Theoretical%20and%20Practical%20Implications.pdf

Trochim, W. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation, 123*, 13–32.

Trochim, W., Urban, J. B., Hargraves, M., Hebbard, C., Buckley, J., Archibald, T., … Burgermaster, M. (2012). *The guide to the systems evaluation protocol*. Ithaca, NY: Cornell Digital Print Services.

Urban, J. B., Archibald, T., Hebbard, C., Burgermaster, M., Barrios, V., & Trochim, W. M. (2015, April). *Results and implications of using the SEP and the mySEP in evaluation planning and implementation for STEM outreach programs*. Paper presented at the Eastern Evaluation Research Society, Absecon, NJ.

Urban, J. B., Hargraves, M., Hebbard, C., Burgermaster, M., & Trochim, W. M. (2011). *Evaluation in the context of lifecycles: 'A place for everything, everything in its place.'* Paper presented at the American Evaluation Association, Anaheim, CA.

Urban, J. B., Hargraves, M., & Trochim, W. (2014). Evolutionary evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning, 45*, 127–139.

Urban, J. B., Linver, M. R., Thompson, J., Davidson, R., & Lorimer, D. (2017). Evaluating youth character education using evolutionary evaluation and the systems evaluation protocol. *Applied Developmental Science*. http://dx.doi.org/10.1080/10888691.2017.1285236

Urban, J. B., & Trochim, W. M. (2009). The role of evaluation in research-practice integration: Working toward the "golden spike." *American Journal of Evaluation, 30*(4), 538–553.

W.K. Kellogg Foundation. (2004). *W.K. Kellogg Foundation logic model development guide*. Battle Creek, MI: Author.

Wandersman, A. (2014). Moving forward with the science and practice of evaluation capacity building (ECB): The why, how, what, and outcomes of ECB. *American Journal of Evaluation, 35*(1), 87–89.

Wikipedia. (2007). Web 2.0. Retrieved December 18, 2007, from http://en.wikipedia.org/wiki/Web_2

Yeager, D. S., Bundick, M. I., & Johnson, R. (2012). The role of future work goal motives in adolescent identity development: A longitudinal mixed-methods investigation. *Contemporary Educational Psychology, 27*(3), 206–217.

Young, G. (1999). A piece of the main: Parts and wholes in ecology and human ecology. *Advances in Human Ecology, 8*, 1–31.