



A Framework For Evaluating Systems Initiatives

Julia Coffman¹
August 2007



I. Background ● ● ●

In April 2007, more than 60 evaluators, funders, state leaders, and national experts came together in Pittsburgh for a symposium on evaluating *systems initiatives* — efforts to build or reform health, education, or human service systems for purposes of improving individual or community well-being. The symposium was sponsored by The Build Initiative, a foundation-funded multi-state effort to ensure that children from birth through age five are safe, healthy, eager to learn, and ready to succeed in school.² The Build Initiative supports states' efforts to build comprehensive and coordinated early childhood systems of programs, policies, and services that work together to achieve positive outcomes for young children and their families.³

Evaluating systems efforts like The Build Initiative in ways that both capture their impact and inform their ongoing development can be a significant challenge. Systems initiatives are complex and notoriously “hard to measure.” They involve multiple programs and players and feature outcomes at multiple levels (individual, family, community, and state). They involve numerous public or private funding streams administered through different agencies and decision-making structures. They require aligning goals and coordinating actions across programs with different political cultures. And, they tackle difficult deep-rooted problems such as gaps in services and outcomes based on race, income, culture, and language. Finally, all efforts to improve systems are long-term efforts that evolve over time in response to the political, economic, and social contexts around them. These many complexities place systems initiatives directly outside of the more familiar and more traditional program evaluation comfort zone. Consequently, less consensus exists about how to assess them.

Still, systems initiative evaluation is not uncharted territory. Systems initiatives have been around for decades and various evaluation approaches have been tried. Of particular note are “theory of change” evaluation approaches that have gained substantial momentum since the mid 1990s when the Aspen Institute’s Roundtable on Comprehensive Community Initiatives for Children and Families introduced them as a promising approach for evaluating complex initiatives. Theories of change are now the cornerstone of many, if not most, systems initiative evaluations.⁴ But while theories of change have added much to evaluation practice in this area, they are not (and did not promise to be) a panacea for all evaluation dilemmas that systems initiatives present. . In practice they have been more a way of describing system elements and systems initiative complexities than an evaluation methodology that spells out initiative assumptions and ways of testing whether they are valid.

Recognizing that systems initiative evaluation has evolved but few efforts have truly examined this emerging field, The Build Initiative determined the time was right for a symposium that would bring together long-time experts in the field with new and diverse voices. Specifically, the symposium was designed to:

- 1) Assess the current state of the systems initiative evaluation field
- 2) Identify lessons learned from previous systems initiative evaluations
- 3) Point to next steps and directions for evaluating systems initiatives.

¹ Julia Coffman is an evaluation consultant based in Alexandria, Virginia. Comments or questions may be directed to jcoffman@evaluationexchange.org. The author gratefully acknowledges the following individuals for their feedback during the paper’s development: Charles Bruner, Susan Hibbard, Anne Kubisch, Marge Petruska, Lisbeth Schorr, Sam Stephens, Michelle Stover-Wright, and Gerrit Westervelt.

² The Build Initiative (www.buildinitiative.org) is a project of the Early Childhood Funders Collaborative, an affiliation of experienced grantmakers at foundations or corporate giving programs with substantial grantmaking portfolios in early childhood care and education.

³ The Heinz Endowments, an Early Childhood Funders Collaborative member, generously supported the symposium and the products produced from it, including this paper.

⁴ Connell, J., Kubisch, A., Schorr, L., & Weiss, C. (1995) (Eds.). *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, DC: The Aspen Institute; Fullbright-Anderson, K., Kubisch, A., & Connell, J. (1998) (Eds.). *New approaches to evaluating community initiatives: Theory, measurement, and analysis*. Washington, DC: The Aspen Institute.



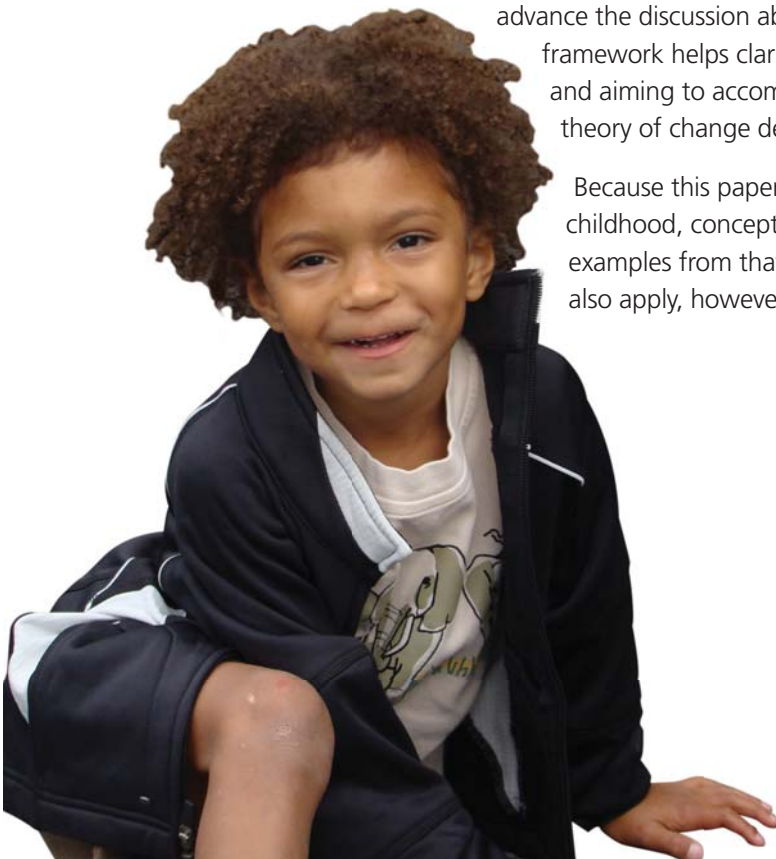
Purpose of this Paper

This paper draws on the symposium's rich debate and discussion. It is not a description of the symposium's proceedings, although it builds on the meeting's content and dynamic discussions. Rather, it offers a framework that attempts to clarify ideas, approaches, and language about evaluating systems initiatives. The paper's purpose is to facilitate reflections about past evaluations and to guide decisions about their future.

The symposium revealed several things that call for such a paper. First, systems initiatives are not homogenous. They attempt to change different aspects of systems and focus on systems at different stages of development. As a result, it can be difficult to have clear conversations about how to evaluate systems initiatives because they can mean different things to different people. Second, because there are many aspects and stages of systems development, no one evaluation approach is sufficient or appropriate for all systems initiatives. Multiple evaluation approaches can be appropriate and useful, with different approaches "fitting" certain initiatives better than others. Until a way of talking about different kinds of initiatives exists, however, it will be difficult to sort out what evaluation approaches to use and when. Finally, many age-old questions still loom large in discussions about evaluating systems initiatives. These include whether experimental designs (or other counterfactuals) are appropriate or even possible in this context; under what conditions systems initiatives should be held accountable for demonstrating individual-level impacts for system beneficiaries; and whether the same evaluation methodologies can meet both the needs of funders and practitioners. While these questions have been asked many times, clear or satisfactory answers have not yet emerged, in part because these questions often are raised at a macro level, where the concept of evaluating complex systems initiatives in their entirety can be overwhelming.

In response to these issues, this paper introduces a framework to help advance the discussion about evaluating systems initiatives. The framework helps clarify what complex systems initiatives are doing and aiming to accomplish, and thereby supports both initiative theory of change development and evaluation planning.

Because this paper grew out of a symposium focused on early childhood, concepts presented throughout are illustrated with examples from that field. The framework and ideas presented also apply, however, to systems initiatives in other fields.



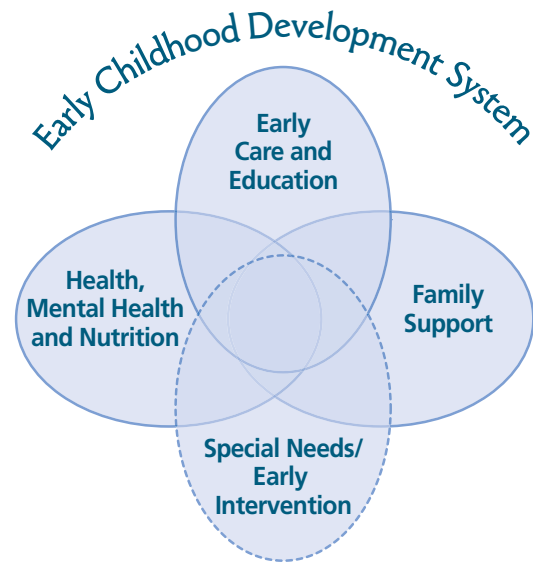
II. Defining a System ● ● ●

“A system is a group of interacting, interrelated, and interdependent components that form a complex and unified whole.”⁵ A system’s overall purpose or goal is achieved through the actions and interactions of its components.

While systems are everywhere—they can be ecological, mechanical, organizational, political, cultural, and so on—the focus here is on health, education, and human service systems. These systems have multiple programs, policies, agencies, or institutions at the national, state, or local level with the common goal of achieving better outcomes for children, youth, adults, or families.

Health, education, and human service systems possess several important characteristics that have implications for evaluation. The remainder of this section describes these characteristics using the example of an *early childhood development system*.

The figure at right, developed by the Early Childhood Systems Working Group⁶, shows four components that the Group has described as part of an early childhood development system—early care and education; family support; health, mental health, and nutrition; and special needs/early intervention.⁷ The idea is that optimal developmental outcomes for young children will be achieved when each component is fully developed and the four components or subsystems connect or align.



Characteristics of this system (as well as other health, education, and human service systems) include:⁸

It contains numerous subsystems.

Each component in the early childhood development system is a “system within a system,” meaning that each has its own set of interacting programs, policies, and strategies that meet certain child development needs. For example, the early care and education component focuses on the early learning opportunities that children need to succeed in school and later on in life. This subsystem includes, for example, family and center-based child care, preschool programs, family literacy programs, and so on. For the system to operate effectively, the subsystems must operate effectively. This includes reaching all children and families for whom the subsystems are designed. Because some subsystems may reach only a fraction of their intended beneficiaries, building them to reach their full potential is an important part of systems development.

It is also part of a larger system. Health, education, and human service systems are open systems, meaning they interact with other systems, have permeable boundaries, and are affected by their external environment. For example, the early childhood development system interacts with and is part of many larger systems. One of these is a larger human service system that in addition to early

⁵ Pegasus Communications. What is systems thinking? *The Systems Thinker*. Retrieved on June 28, 2007 from <http://www.thesystemsthinker.com/systemsthinkinglearn.html>.

⁶ A collaboration of 12 national organizations focused on early childhood issues in the United States. Groups include the Alliance for Early Childhood Finance, The Build Initiative, The Children’s Project, Center for Law and Social Policy, Council of Chief State School Officers, National Center for Children in Poverty, National Child Care Information Center, National Council of State Legislatures, National Governor’s Association Center for Best Practices, Smart Start National Technical Assistance Center, State Early Childhood Policy Technical Assistance Network, and Zero to Three.

⁷ This oval has a dotted line to signify that better integration of the Special Needs/Early Intervention system into the other three components is a goal, but is not yet a reality in part because of funding streams, requirements, and state structures.

⁸ Informed by Pegasus Communications. What is systems thinking? *The Systems Thinker*. Retrieved on June 28, 2007 from www.thesystemsthinker.com/systemsthinkinglearn.html.

childhood includes services for older children and adults. Human service systems also, for example, cover economic assistance and income supports, employment, food stamps, disability, rehabilitation, and senior services.

Interconnections are essential for optimal results. A system typically has a goal or function that is best achieved when its components function together as a whole. More colloquially, with systems, the whole is greater than the sum of the parts. Applied to health, education, or human service systems, this means that a connected web of services and programs will result in better outcomes for individuals than if those services and programs are not connected. In the early childhood development system, for example, while the four components are important in and of themselves, a long tradition of child development research shows that these contexts have reciprocal influences on each other and on children.⁹ Being strategic about the ways in which they connect so that their collective strengths can be leveraged, will be more effective in improving child outcomes than if those components operate independently. A simple example is that children learn best when they receive a nutritional breakfast. Therefore connecting early care and education programs with food and nutrition programs should result in better developmental outcomes. This “system of systems” concept is illustrated where the components in the system figure’s center (on page 3) overlap. On a more complex level, some families need multiple services and supports involving different subsystems, such as child care, family support, and parenting education. Services will have a lesser impact if their philosophies and approaches for working with parents are not aligned. For example, families with multiple case managers need consistent directives so that these supports do not overwhelm families rather than help them. Coordination and alignment allows subsystems to achieve their individual objectives while contributing to larger system goals.

It is a “loosely coupled” system. Health, education, and human service systems are loosely coupled, meaning that actions taken in one part of the system may have little or no consequence for other system parts, and the effects that do result can be unpredictable. Additionally, a loosely coupled system’s parts tend to be autonomous or self-directed, and the incentives for them to integrate can be weak.

For the early childhood development system, for example, this means that while early care and education and family support are in the same system, there are no serious consequences to the system if they are not closely connected (there are, however, consequences for children). Even though interconnections are essential for optimal child outcomes and that by itself should be an incentive to connect, the system will not “break down” if they do not. Benefits of loosely coupled systems, on the other hand, are that they allow for creative changes and adaptations to occur within individual subsystems without throwing the whole system off balance. Similarly, individual subsystems can fail or be incomplete without damaging the whole system.¹⁰

These characteristics demonstrate some of the complexities that systems bring to evaluation. For example, because systems are loosely coupled, it is hard to predict or trace the location and strength of results after improvements are introduced. This makes evaluation, which prefers more predictable and controlled conditions, difficult at best. These characteristics are important to keep in mind during the more specific discussion about evaluating systems initiatives in the paper’s later sections.

⁹ E.g., Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513-530.

¹⁰ Weick, K. (1976). Educational organizations as loosely coupled systems [Electronic version]. *Administrative Science Quarterly*, 21, 1-9.



III. Defining Systems Initiatives: Five Focus Areas ● ● ●

Few systems function perfectly. For example, components that would greatly improve their functioning may be missing, or the system's components may be disconnected and working toward inconsistent goals. This is especially true with human service systems, where programs are often disconnected or "siloed" and consequently feature duplications or gaps in services.

Systems initiatives are organized efforts to improve a system and its impacts. They can be publicly or privately funded or a combination of the two. All systems initiatives, however, are not the same. They may have different goals and objectives or work on different aspects or stages of systems development. It is important to understand their distinctions because, as described later, different systems initiatives call for different evaluation approaches.

Systems initiatives sometimes have different labels, such as systems building, systems change, or systems reform. While systems *building* initiatives often refer to efforts to create the policies, funding, programs, or services that make up the system, and systems *Change* or *reform* efforts tend to take place on the "front lines" of practice, in fact these labels are used inconsistently and therefore do little to distinguish their differences. Instead, systems initiatives are best understood by their focus or by the parts of the system they are trying to improve.

Specifically, a systems initiative might focus on one or more of these five areas.

Context—Improving the political environment that surrounds the system so it produces the policy and funding changes needed to create and sustain it

Components—Establishing high-performance programs and services within the system that produce results for system beneficiaries

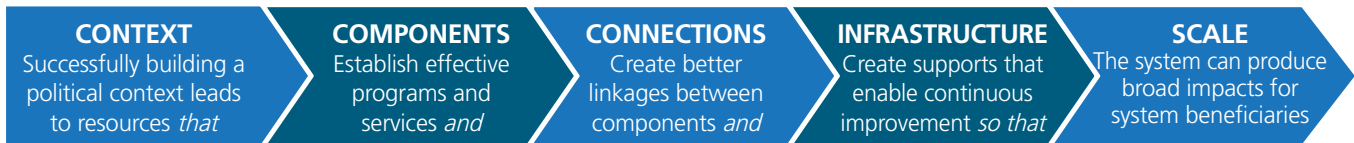
Connections—Creating strong and effective linkages across system components that further improve results for system beneficiaries

Infrastructure—Developing the supports systems need to function effectively and with quality

Scale—Ensuring a comprehensive system is available to as many people as possible so it produces broad and inclusive results for system beneficiaries.



These five areas comprise the aspects of a system that, if developed or advanced, can produce broad impacts for the system's intended beneficiaries. The figure below illustrates the basic logic of how these areas work together to produce systems-level impact.



Note that systems initiatives do not have to focus on all five areas. Some may focus only on one or two. Most systems initiatives, however, focus on more than one area, and many focus on four or five.

Also, while the figure implies a linear sequence, in fact, initiatives can focus on multiple areas simultaneously. They do not, however, typically place an equal emphasis on all focus areas at once. Some areas receive more attention than others at any given point in time, depending on where the system's needs are greatest and the opportunities that are available.

This emphasizes again that while systems initiatives share similarities in the kinds of things they focus on; each initiative is constructed differently. Before evaluations are designed, then, it is important to have a clear map of what initiatives are doing relative to what they are trying to achieve, and how the change process is expected to occur. In other words, the evaluation process should begin with a theory of change.



IV. Developing Systems Initiative Theories of Change ● ● ●

Theories of change are now part of mainstream evaluation practice; almost every systems initiative evaluation is now based on one. Theories of change illustrate the pathways by which change is expected to occur and the role that initiatives play in producing that change. They show how initiative strategies or activities connect to interim outcomes that then set the stage for long-range impacts. Because theories of change enable evaluation stakeholders to clarify what initiatives are doing in relation to their intended outcomes and impacts, they typically are developed early on to inform both initiative strategy development and evaluation design.

As theories of change have become ubiquitous, they have taken on many forms. Some are depicted visually while others are narratives. Some are linear while others are circular. Some are broad while others are quite specific. The point here is not to prescribe a specific formula for how a theory of change must look. Rather, it is to say that the concept that underlies a theory of change—making explicit stakeholder notions about the relationships between an initiative’s strategies, interim outcomes, and long-term impacts, and producing testable assumptions regarding those relationships—is always useful, especially for complex initiatives where it can be difficult to understand the many strategies in play.

A Theory of Change Menu for Systems Initiatives

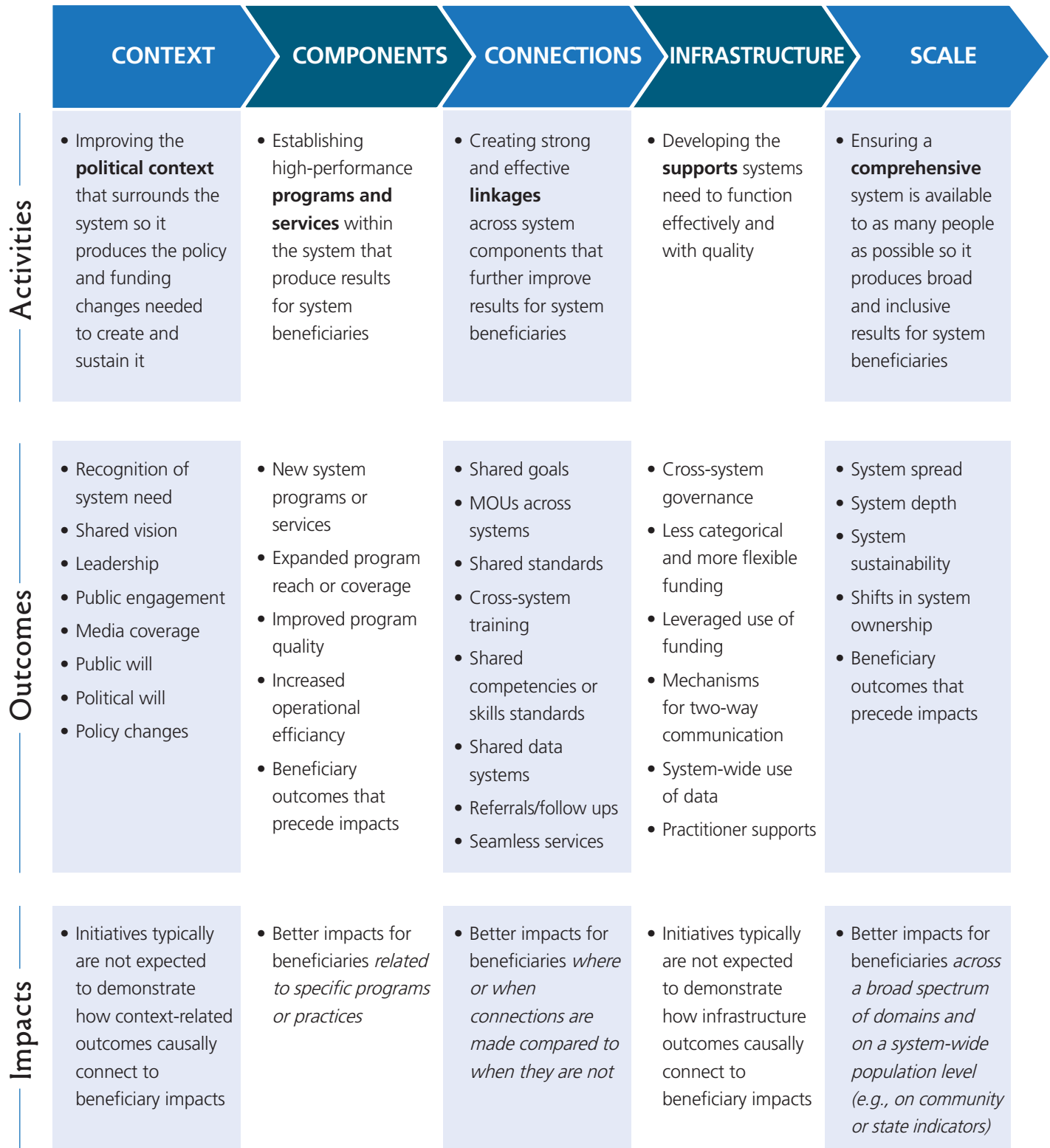
The figure on the next page was developed to aid theory of change development for systems initiatives. It is constructed around the five systems initiative focus areas. For each focus area, the figure offers a broad description of initiative activities, possible outcomes of those activities, and what those outcomes, if achieved, may produce in terms of impacts for system beneficiaries. The actual outcomes and impacts selected for a theory of change will depend on the systems initiative at hand. For example, an initiative focused on both a system’s context and its infrastructure would have a theory of change featuring content from both of those columns.

The figure is simply a menu. Actual theories of change should detail how initiative strategies will connect to and produce their intended outcomes and impacts. For an initiative focused on context, for example, it is important to describe or illustrate how outcomes such as public will or leadership will be built. In other words, an initiative with multiple areas of focus might have both a broad theory of change that includes outcomes and impacts across focus areas, and a more detailed version that breaks out and specifies the theories of change within each focus area.

The activities, outcomes, and impacts associated with each focus area are outlined in more detail below and then illustrated using examples of real-life initiatives. Again, note that systems initiatives typically focus on more than one of these areas simultaneously.



A Theory of Change Menu for Systems Initiative ● ● ●



1 Context

Initiatives focused on context attempt to change the political environment that surrounds and affects a system's development and ultimate success.

Activities

Most human service systems require substantial public investments. They cannot be built entirely on private resources; nor can they be scaled up or sustained without public funding.¹¹ Securing these public investments requires political will. Some systems initiatives therefore focus on building that political will and on the elements that influence it. Activities may include, for example, public education and awareness building, leadership development, media advocacy, grassroots or grassroots organizing, coalition building, or community mobilization.

A focus on context tends to come earlier in a system's developmental trajectory, as it can help pave the way for the system's construction. At the same time, systems also require that context be a continuous focus, as sustaining support for systems requires constant vigilance in a regularly shifting political environment.

Outcomes

It is useful to ground theories of change for systems initiatives that focus on context in theories about the policy process, as these theories point to the factors that influence the policy environment and signal what outcomes to measure. For example, basic agenda setting theory says that the policy agenda is influenced by what the public thinks and cares about. Public opinion, in turn, is influenced by what the media reports and how they frame issues. It follows, then, that initiatives wanting to get a particular systems issue on the policy agenda may use the media as a vehicle for doing so. An evaluation to track the initiative's success in influencing the policy agenda would then track its progress toward this goal by looking at the shorter-term outcomes of both media attention and public opinion.¹²

For initiatives focused on context, an array of outcomes is possible. Those outcomes include, but are not limited to:

- Increases in issue awareness or salience
- Public engagement or mobilization
- New advocates or champions
- Issue recognition or reframing
- Public will
- Political will
- New or more diversified funding
- Policy changes

¹¹ Bruner, C. (2004). *Toward a theory of change for the Build Initiative: A discussion paper*. Retrieved on June 27, 2007 from <http://www.buildinitiative.org/docs/TowardTheoryofChange.doc>.

¹² It also may be possible to produce change in context through other approaches, including interest group pressure and advocacy independent from broad public opinion. See, for example, Bruner, C. (2004). *Beyond the usual suspects*. Des Moines, IA: The State Early Childhood Policy Technical Assistance Network (SECPTAN).



Impacts

While better services and outcomes for children, families, communities, etc. are always the end goal, systems initiatives typically are not expected to demonstrate how context-related outcomes causally connect to beneficiary impacts. Evaluations tend to focus instead on outcomes that are “closer” to the initiative’s work (like the ones described above). Faced with a choice of how to invest limited evaluation resources, it usually makes more sense to assess the outcomes that more directly relate to initiative efforts as they deliver information that usefully informs the evolving initiative strategy.

Examples

The Build Initiative is an example of an initiative with a major focus on political context.¹³ Build states attempt to gain support for state-level early childhood development systems in the political environment that surrounds and profoundly affects them. In fact, the Build Initiative theory of change includes eight critical elements required to produce change in a political context: 1) recognition of need, 2) a shared early learning vision, 3) political leadership, 4) capacity and expertise, 5) programs, actions, and policy successes, 6) public awareness and support, 7) mobilization and advocacy, and 8) alignment and readiness.¹⁴

Another systems initiative focused on political context is the Urban Health Initiative funded by the Robert Wood Johnson Foundation. This ten-year initiative used a systems change strategy to improve the health and safety of children and youth in select cities across the nation. To do this, it focused on shifting policies and funding so that child well-being would move to the top of the policy agenda. Making these shifts required strategies to build both public and political will.¹⁵

2 Components

Systems initiatives focused on components concentrate on putting in place high-quality and high-performance programs, services, or interventions for the system’s intended beneficiaries.

Activities

While a system is made up of interconnected parts and those connections are critical, sometimes a system’s problem is not so much that it lacks connections, but that it lacks the parts to connect in the first place. Although a system must contain more than independently operating programs, ensuring that a sufficient and comprehensive supply of quality and affordable programs and services are in place should not be overlooked as a potentially important part of systems work.¹⁶

Outcomes

Outcomes for initiatives focused on system components will depend on the program or intervention at hand. They may relate to the program itself (e.g., participation, quality, or operational efficiency) or to participant-level outcomes that precede impact (e.g., improvements in child nutrition as a precursor to improvements in early learning outcomes). Outcomes in this area may include:

¹³ As shown later, the Build Initiative actually includes a simultaneous focus on four of the five focus areas.

¹⁴ Ibid.

¹⁵ Hackman, S. (2005). Fielding large-scale community change initiatives: Key lessons from a decade of systems change [Electronic version]. Cambridge, MA: The Urban Seminar Series on Children’s Health & Safety.

¹⁶ Bruner, C. (2004). *Toward a theory of change for the Build Initiative: A discussion paper*. Retrieved on June 27, 2007 from <http://www.buildinitiative.org/docs/TowardATheoryOfChange.doc>.

- New programs or services developed within the system
- Expanded individual program reach or coverage
- Improved program quality
- Increased program operational efficiency
- Beneficiary outcomes that precede impact

Impacts

Systems initiatives focused on individual programs may be expected to demonstrate measurable impacts for program participants. For example, a nutrition program would expect better health results for participants. In contrast to evaluations of more comprehensive systems where impacts are expected across a broad array of domains, expected impacts here are specific to the program or intervention being evaluated.

Example

In addition to its focus on context, The Build Initiative focuses on system components. While attempting to make improvements in the political environment that enable early childhood development systems to be built, Build states also work to improve the individual subsystems in them. This step is essential to ensuring that children and families have access to the *comprehensive* set of programs and services they need to be successful. For example, some states are working on enhancing their early and education subsystem by making high-quality preschool programs available for all three- and four-year-olds. Others are concentrating on creating infant mental health programs and services. Still others are developing strategies to strengthen the quality of programs that currently exist. While all of the states have a broader system building vision and an accompanying strategy focused primarily on political context, they take advantage of opportunities to advance the development of specific subsystems and see this work as very much a part of their overall systems efforts.

3 Connections

Systems initiatives that concentrate on connections focus on what makes a system a *system*—the integration, linkages, and alignment between its parts.

Activities

Initiatives focused on this area attempt to connect subsystems or programs in meaningful ways. On the frontlines, those connections may include, for example, programs across or within subsystems that share professional development, staffing, facilities, technology and communication, data collection, or funding. At the administrative level, such connections may include, for example, aligned eligibility requirements and enrollment processes, streamlined reporting procedures, coordinated case management, and established protocols and memoranda of understanding for referrals across subsystems. These kinds of connections ensure that, when necessary, beneficiary needs identified in one subsystem can be referred to and managed by another.

Outcomes

Outcomes here relate to connections themselves in the form of increased system coordination, alignment, integration, or linkages. The challenge is operationalizing these concepts so they are measurable and relevant to the systems initiative at hand. Ways of operationalizing connections include, for example:

- Coordinated eligibility assessments
- Referrals occurring from one program to another
- Activities to ease within-system transitions
- Joint planning across system components
- Cross-system competencies or skills standards
- Cross-system training
- Shared data systems for tracking individuals
- Memoranda of agreements between system components

Impacts

Systems initiatives focused on connections may be expected to demonstrate impacts for system beneficiaries. Better results for beneficiaries are expected when the system is connected compared to when it is not. The specific individual-level impacts measured will depend on the level at which connections are expected. If many connections are expected across the system, then impacts can be expected across a wide array of domains (e.g., health, literacy, mental health, etc.). If they are expected within or between particular subsystems, then expected impacts should be more specific to those areas.

Example

The W.K. Kellogg Foundation's SPARK Initiative focuses on connections. It funds partnerships in seven states and the District of Columbia to ensure that children are ready for school and that schools are ready for children. SPARK (Supporting Partnerships to Assure Ready Kids) sites have adopted alignment models that take into account schools' relationships with the community and with the early care and education system. They support seamless transitions from early childhood through elementary school by aligning these systems with health and other service systems. SPARK is demonstrating that it is possible for the early care and education system to connect with nontraditional partners like public health on behalf of kids.¹⁷



¹⁷ W.K. Kellogg Foundation. SPARK overview. Retrieved on June 29, 2007 from <http://www.wkkf.org/default.aspx?tabid=75&CID=168&NID=61&LanguageID=0>.

Infrastructure

Sometimes a system's main problems or needs do not exist as much at the actual point of service as at the level of governance, financing, or other critical supports that challenge their functioning or existence.¹⁸

Activities

Systems initiatives focused on infrastructure make changes that facilitate a system's development and functioning. Infrastructure development helps to ensure that systems have the supports they need to function effectively and with quality.¹⁹

Outcomes

Outcomes for systems initiatives focused on infrastructure depend on the kinds of supports the initiative aims to establish or enhance. For example, infrastructure outcomes may include:

- Governance entities that oversee and coordinate subsystems
- Funding streams that are less categorical
- Leveraged use of funding
- Research and monitoring that encourages the cross-system use of data
- System-level or aligned standards
- Professional development structures that can include pre-service and in-service education to ensure an appropriately skilled systems workforce
- Practitioner training and technical assistance to support on-the-ground systems development.

Outcomes also may include interim benchmarks on the way to achieving longer-term infrastructure outcomes. For example, if the goal is establishing a new governance structure and this goal requires a policy change, then the evaluation might examine shorter-term outcomes on the way to reaching that goal, such as policymaker recognition that a new structure is needed and political will for making that change happen. As this example shows, systems initiatives that focus on infrastructure tend to concentrate simultaneously on other focus areas. In particular, opportunities to achieve infrastructure outcomes often are the result of changes that first occur in the political context.

Impacts

Like systems initiatives focused on context, initiatives typically do not have to causally link infrastructure outcomes to impacts for beneficiaries. Again, the kinds of outcomes listed above tend to be the evaluation's focus because they are "closer" to what the initiative is actually doing.

Example

The [North Carolina Smart Start Technical Assistance Center](#) is a foundation-funded initiative to support states in their early childhood systems building efforts.²⁰ The Center was established to share lessons from North Carolina's experiences with its own early childhood systems effort called Smart Start.

¹⁸ Gonzalez, R., & Gardner, S. (2003). *Systems change and school readiness* [Electronic version]. UCLA Center for Healthier Children, Families, and Communities; Chynoweth, J., Philliber, S., & Oakley, M. (2000). *Systems change in California: Progress at the Millenium*. Sacramento, CA: The Foundation Consortium.

¹⁹ Based on the "core elements of an early childhood development system" developed by the Early Childhood Systems Working Group.

²⁰ Funders include the W.K. Kellogg Foundation, David and Lucile Packard Foundation, Carnegie Corporation of New York, Triad Foundation, and Heron Foundation.

Fourteen years ago, a favorable political environment and gubernatorial leadership in North Carolina led to Smart Start, a statewide initiative that provided substantial new funding for early childhood systems building. Smart Start's challenge was to use those resources effectively. In response, the state developed an infrastructure that included governance structures, evaluation and research components, and technical assistance to communities based on best practices. That infrastructure, also supported by the private sector, created a continued favorable political environment for growth and development, as it provided policymakers and leaders with confidence that systems development was both deliberate and accountable.

The Center was founded on the notion that North Carolina's deep experience and lessons learned could be usefully applied in other states where systems work was in an earlier stage of development. Center staff members use technical assistance, coaching, site visits, research, conferences, and other resources to support states and share lessons. They tend to concentrate on states' infrastructure and capacity needs, where lessons are particularly transferable (e.g., sharing how different states have structured their governance structures or creative ways in which to use federal or state funding streams to support systems work).

The Annie E. Casey Foundation's Casey Strategic Consulting Group is another example of an initiative focused on infrastructure to support human service system reform. Founded in 2001, the Strategic Consulting Group works directly and on-site with state agencies to strengthen their infrastructure for systems, focusing on, for example, agency management, internal operations, and frontline practices. The Group has adapted private sector management consulting practices for public sector purposes, and uses a team-based approach combined with strategic analysis and rigorous research to assess system performance, identify gaps, and develop solutions. The Group also brings to its work the Casey Foundation's resources in the form of funding (consulting services are free for clients selected), deep expertise on systems reform, and a respected reputation in the public sector.²¹ The Strategic Consulting Group selects partners with favorable political contexts—those that show readiness for “non-incremental” change within major human service systems.

5 Scale

Systems initiatives focused on scale ensure that a comprehensive and quality system is available to as many of its intended beneficiaries as possible.

Activities

Initiatives may attempt to scale up a system by increasing system services and the numbers of beneficiaries served; ensuring program integrity and service quality is high enough to be consequential; and making sure systems are sustainable.²²

Initiatives attempting to scale up a system usually require a high level of funding. This funding can come from both public and private investments, but more often than not the lion's share must come from public sources, especially if the goal is to scale up the system statewide.

²¹ Annie E. Casey Foundation. Casey strategic consulting. Download on June 30, 2007 from <http://www.aecf.org/MajorInitiatives/CaseyStrategicConsulting.aspx>.

²² Coburn, C. (2003). Rethinking scale: Moving beyond the numbers to deep and lasting change [Electronic version]. *Educational Researcher*, 32 (6), 3-12.

Outcomes

Outcomes for systems initiatives focused on scale depend on how scale is defined. Scale definitions may include:

- System spread
- System depth
- System sustainability
- Shifts in system ownership

For health, education, or human service systems, traditional definitions of scale concentrate on the system's *spread*, or increases in the number of programs and people served. But this definition alone is limited; it says little about how scale up has affected the services people receive. The concept of spread should expand beyond simply "upping the numbers" to include the spread of ideas, beliefs, values, and principles that support bringing systems to scale. *Depth*, or the nature and quality of change that takes place at the practice level, is another important dimension of scale, as is *sustainability* or whether the system can be maintained over time. Finally, the definition of scale can include a *shift in ownership*, meaning that a broad array of stakeholders, especially those on the frontlines, assume responsibility for maintaining the scaled-up system.²³

Outcomes also may include various benchmarks for longitudinal impact measures (e.g., expecting incremental improvements in low birth weight statistics over time), or the individual-level outcomes that precede impacts (e.g., more mothers receiving prenatal care contributing to better birth weight statistics).

Impacts

Systems initiatives that concentrate on scale generally are expected to demonstrate impacts for system beneficiaries. Because the system is, by definition, comprehensive, these impacts should exist across a wide range of domains.

Examples

The First 5 California School Readiness Initiative is an example of a systems initiative focused on scale. Supported by funding from a tobacco tax passed in 1998, it is the largest early childhood and school readiness initiative in the nation (over \$450 million has been allocated since 2002). The initiative provides funding to all of California's counties for the development and functioning of community-based early childhood systems that provide and coordinate high-quality programs for children ages birth to five in areas that include early care and education, parenting and family support, health and social services, and school capacity. This initiative falls into the scale category because it provides funding for movement toward comprehensive early childhood development systems statewide. These systems are expected to be comprehensive and accessible enough to produce positive changes in a broad spectrum of child and family outcomes that relate to school readiness.²⁴

²³ Coburn, C. (2003). Rethinking scale: Moving beyond the numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.

²⁴ First 5 California. First 5 California School Readiness Initiative. Retrieved on June 30, 2007 from <http://www.ccfcc.ca.gov/SchoolReady1.htm>.

North Carolina's Smart Start is another example in the scale category. This now 14-year-old initiative was established to build a comprehensive and coordinated early childhood system statewide for purposes of improving school readiness outcomes. The initiative offers children birth to age five a range of services with the goal of enabling children to enter school healthy and ready for success. Smart Start is funded through public dollars drawn from the state's general fund that are then directed to local partnerships in all of North Carolina's 100 counties (the state appropriation was \$204 million in fiscal year 2007; funding peaked at \$231 million in fiscal year 2000).²⁵ While the majority of Smart Start funding is spent on child care or child care-related services (e.g., subsidies), about one-third of local partnership dollars must address other issues affecting young children and their families, such as family support and health care.²⁶

V. Designing Systems Initiative Evaluations ● ● ●

As the last section makes clear, all systems initiatives are not the same. It follows then that their evaluations should not be the same. Systems initiative evaluations should be tailored to their unique theories of change, assessing the outcomes and impacts connected to the parts of the system they are attempting to change.

The figure on the next page was developed to guide evaluation planning decisions for systems initiatives. Like the theory of change menu, it is organized around the five focus areas and offers ideas about appropriate evaluation choices for initiatives that incorporate each.

Each focus area features two evaluation questions that generally address:

- Did the initiative do what it said it would do (in that focus area)?
- Did the initiative produce the expected results (for that focus area)?

The figure also features possible evaluation methodologies to address those questions. These lists of designs and methods are not exhaustive or prescriptive. They offer ideas, but do not spell out how evaluations should look. Those decisions must be based on a careful consideration of the specific initiative in question and what it is designed to do.

Again, it is important to keep in mind that most systems initiatives are hybrids and focus on more than one area. Consequently, multiple evaluation questions will apply and evaluations should feature a mix of design elements to address those questions. Additionally, evaluations should identify the areas in which the initiative places the most emphasis, and plan for similar levels of evaluation emphasis in these same areas.



²⁵ Center for Law and Social Policy. North Carolina Smart Start. Retrieved on June 30, 2007 from <http://www.clasp.org/ChildCareAndEarlyEducation/map030707nc1.htm>.

²⁶ Coffman, J., Stover Wright, M., & Bruner, C. (2006). *Beyond parallel play: Emerging state and community planning roles in building early learning*. Des Moines, IA: State Early Childhood Policy Technical Assistance Network

An Evaluation Design Menu for Systems Initiatives ● ● ●

	CONTEXT	COMPONENTS	CONNECTIONS	INFRASTRUCTURE	SCALE
Questions	<ol style="list-style-type: none"> Has the initiative changed the political environment through its activities? Has the initiative produced changes to investment, policy, or practice that will enable changes in components, connections, infrastructure, or scale? 	<ol style="list-style-type: none"> Did the initiative design and implement system components as intended? Did the components produce their intended impacts for beneficiaries? 	<ol style="list-style-type: none"> Did the initiative design and implement connections and linkages as intended? Did the connections and linkages produce their intended impacts? 	<ol style="list-style-type: none"> Did the initiative establish infrastructure or supports that are consistent with its objectives? Did the Infrastructure or supports achieve their objectives for effectiveness, sustainability, and quality? 	<ol style="list-style-type: none"> Did the initiative enable system scale up with quality and fidelity? Did scale up result in broad impacts for beneficiaries at a system-wide population level?
Methodologies	<ul style="list-style-type: none"> Theory of change evolution Case studies Public polling Policy tracking Key informant surveys Coalition analysis Policy/bellwether interviews Media tracking 	<ul style="list-style-type: none"> Program evaluation methodologies (including experimental/quasi-experimental) Program monitoring Quality assessments Efficiency analyses Customer surveys 	<ul style="list-style-type: none"> Program evaluation methodologies (including experimental/quasi-experimental) System mapping Network analysis Customer surveys 	<ul style="list-style-type: none"> Theory of change evaluation Case studies Performance audits Management information systems Practitioner data collection 	<ul style="list-style-type: none"> Population-based demographic and service analysis Program evaluation methodologies (including experimental/quasi-experimental) System/program monitoring Results-based accountability

Evaluation questions, designs, and methods can be “mixed and matched” as appropriate. For example, a quasi-experimental design may co-exist with a theory of change approach, or case studies may be used alongside a results-based accountability approach. Initiative evaluations can incorporate multiple designs and a wide range of data collection methods.

The remainder of this section describes evaluation options for each focus area in more detail, discussing relevant evaluation questions and methodologies. In addition, it offers information about the evaluations of the real-life initiatives described earlier.

1 Evaluating Context

Systems initiatives focused on context attempt to affect the political environment so it better supports systems’ development and success.

Evaluation Questions

The key evaluation questions for these initiatives are:

- 1) Has the initiative changed the political environment through its activities?
- 2) Has the initiative produced changes to investment, policy, or practice that will enable changes in components, connections, infrastructure, or scale?

Evaluation Methodologies

Initiatives with a context focus generally use a theory of change approach to evaluation. Once a theory is developed for how outcomes in the political context will be achieved, evaluators seek empirical evidence that the theory’s components are in place and that the theorized links between them exist. In other words, they compare the theory with actual experience.²⁷ Evaluators “construct methods for data collection and analysis to...examine the extent to which program [or initiative] theories hold. The evaluation should show which of the assumptions underlying the program break down, where they break down, and which of the several theories underlying the program are best supported by the evidence.”²⁸

Mostly, evaluations that use this approach will not attempt to determine whether the theory’s components are causally linked. Rather, they assess the components separately, using multiple and triangulated methods where possible, and then use both data and evaluative judgment to determine whether a *plausible and defensible* case be made that the theory of change worked as anticipated and that the initiative had its intended effects. As the Urban Health Initiative evaluation example below demonstrates, however, it is possible to incorporate design elements like comparison groups or other counterfactuals into a theory of change approach to strengthen confidence in interpretations about the links between theory of change components.

²⁷ Connell, J., Kubisch, A., Schorr, L., & Weiss, C. (1995) (Eds.). *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, DC: The Aspen Institute.

²⁸ Weiss, C. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families (p.67). In Connell, J., Kubisch, A., Schorr, L., & Weiss, C. (Eds.) *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, DC: The Aspen Institute.

Because they interact with the policy process, initiatives focused on context share many things in common with advocacy and policy change efforts in general. Both may include similar strategies such as coalition building, communication campaigns, grassroots organizing, and media advocacy. And both are “hard to measure” because they evolve over time and their activities and outcomes can shift quickly.²⁹

These similarities in strategy and purpose allow systems initiatives focused on context to draw from the now growing body of work on advocacy and policy change evaluation to identify useful evaluation approaches.³⁰ For example, evaluation methods that are applicable to both advocacy efforts and systems initiatives may include public polling, media tracking, policy tracking, policymaker or bellwether interviews, or intense-period debriefs.³¹

Examples

The Build Initiative evaluation uses a theory of change approach. The initiative has three theories of change that focus on 1) what an early childhood system must include to produce results; 2) the actions or strategies needed to build an early childhood system, and 3) what special role outside Build supports can provide to catalyze change.³² Since Build began in 2002, the evaluation has focused on the second theory of change. In recent years, the evaluation has attempted to plausibly trace state-level policy and funding changes back to Build Initiative activities.

The Child and Family Policy Center leads the evaluation and partners with local evaluators to collect data in the Build states. Local evaluation partners document theory of change components using a case study approach that employs methods they determine are most relevant in their state, such as interviews, surveys, document review, and participant observation. Evaluators use data from these methods to critically examine the state’s progress around the Build theory of change and to determine whether system-related results can be linked back to Build Initiative activities. The Child and Family Policy Center then produces an annual report on overall initiative progress that includes a cross-state analysis of results.

Another example comes from the Urban Health Initiative evaluation conducted by New York University’s Center for Health and Public Service Research. In a unique design choice, evaluators integrated a theory of change approach with a quasi-experimental comparison group design. Evaluators identified 10 non-initiative cities to compare with initiative cities on outcome and impact measures, including leadership, collaboration, and the use of data. Like other theory of change evaluations, evaluators compared program theory and experience; but they believed they could strengthen their approach by integrating a comparison group into the design to rule out alternative explanations for evaluation findings.³³

²⁹ Harvard Family Research Project (2007). Advocacy and Policy Change. *The Evaluation Exchange*, 13(1). Cambridge, MA: Author.

³⁰ Ibid.

³¹ For a description of the bellwether methodology see Blair, E. (2007). Evaluating an issue’s position on the policy agenda: The bellwether methodology. *The Evaluation Exchange*, 13(1), 29. For a description of the intense-period debrief see Bagnell Stuart, J. (2007). Necessity leads to innovative evaluation approach and practice. *The Evaluation Exchange*, 13(1), 10-11.

³² Bruner, C. (2004). *Toward a theory of change for the Build Initiative: A discussion paper*. Retrieved on June 27, 2007 from <http://www.buildinitiative.org/docs/TowardTheoryofChange.doc>.

³³ Weitzman, B.C., Silver, D. & Dillman, K. (2002). Integrating a comparison group design into a theory of change evaluation: The case of the Urban Health Initiative. *American Journal of Evaluation*, 23(4), 371-385.

2 Evaluating Components

Initiatives that concentrate on a system's components attempt to improve the system by shoring up its individual subsystems, programs, or interventions. For example, these initiatives may pilot new programs, expand access to particular programs or services, or introduce quality improvement initiatives.

Evaluation Questions

Evaluations of initiatives focused on components share much in common with traditional program evaluations, as they both assess individual programs or interventions. Also like program evaluations, initiative evaluations in this area address questions in two main areas—program implementation and program impacts. Key questions include:

- 1) Did the initiative design and implement system components as intended?
- 2) Did the components produce their intended impacts for beneficiaries?

Evaluation Methodologies

Again, because the focus is on individual programs or interventions, evaluations here can borrow from traditional program evaluation approaches which feature the systematic application of social science research designs and methods to assess the implementation and effectiveness of programs or interventions.

Evaluations that examine implementation use some form of program monitoring or process evaluation. Program monitoring addresses questions about 1) the extent to which the program is reaching its target population, 2) whether program delivery matches design expectations, and 3) what resources have been used to deliver the program.³⁴ Program monitoring often goes hand in hand with impact assessments (see below) as monitoring addresses questions about *why* a program was or was not effective. A wide array of both quantitative and qualitative methods can be used for program monitoring, such as observations, participant surveys or focus groups, staff member interviews, and document or record reviews.

Evaluations that examine questions about program impacts may use experimental or quasi-experimental designs that employ a range of possible quantitative or qualitative methods, although quantitative data generally prevail in impact assessments. These designs assign (randomly or non-randomly) individuals to participant and non-participant groups and then compare those groups using, for example, repeated measurements. Experimental or quasi-experimental designs generally provide the most definitive attributions of causality and remain the program evaluation “gold standard.” They are expensive to construct and implement, however, and random assignment may not be appropriate for programs that feature enrollment inclusivity and openness because they violate the program's design.

³⁴ Rossi, P., & Freeman, H. (1993). *Evaluation: A Systematic Approach*. Newbury Park, CA: Sage Publications.

Examples

The Build Initiative in Pennsylvania offers an example of an evaluation for a program introduced to improve one part of the state's early childhood system. Keystone STARS is Pennsylvania's quality rating system to help center-based, family-based, and group home child care programs improve their quality (making it part of the early care and education subsystem). It gives programs a quality "star rating" from one to four stars and then rewards programs that receive higher ratings. The first star is relatively easy to attain and each level up is increasingly difficult.

Keystone STARS was piloted in 2002 and scaled up statewide in 2004. In 2006, the Pennsylvania Department of Public Welfare Office of Child Development, University of Pittsburgh Office of Child Development, and Pennsylvania State University Prevention Research Center evaluated the rating system to determine if it was improving quality in participating child care sites.

The evaluation used a quasi-experimental design that compared randomly-selected participating and non-participating child care programs. It examined the rating system's impact on child care quality, but did not assess its impact on children. Data were collected from 572 child care sites (356 child care centers, 81 group child day care homes, and 135 family child day care homes). Programs were assessed using standard quality assessments—the Early Childhood Environment Rating Scale—Revised (ECERS-R) for center-based programs and the Family Day Care Rating Scale (FDCRS) for home-based programs. Ratings were then compared to baseline data collected in 2002. Analyses as a whole revealed that Keystone STARS was in fact helping child care sites improve their overall quality, and was helping to reverse the negative trend in child care quality that was evident during the prior decade.³⁵

3 Evaluating Connections

Systems initiatives focused on connections try to make sure the whole system is greater than the sum of its parts by improving the relationships between the system's components.

Evaluation Questions

Initiative evaluations in this area tend to focus on two main questions:

- 1) Did the initiative design and implement connections and linkages as intended?
- 2) Did the connections and linkages produce their intended impacts?

Evaluation Methodologies

If the evaluation is assessing the first question and determining whether connections exist within the system, several methodological choices are available. For example, social network analysis is one option. This method explores whether connections or relationships exist, as well as their nature and strength. It identifies the "nodes" (people, groups, or institutions) that make up the network or system, and then examines the relationships between them using mathematical algorithms. Results are depicted visually to reveal the network's connections.³⁶ Pre and post analyses can determine whether the network and its connections look different over time.

³⁵ Barnard, W., Etheridge Smith, W., Fiene, R., & Swanson, K. (2006). *Evaluation of Pennsylvania's Keystone STARS Quality Rating System in Child Care Settings* [Electronic Version]. Harrisburg, PA: Pennsylvania Office of Child Development.

³⁶ Durland, M., & Fredericks, K. (Eds.). (2005). *New directions for evaluation: Social network analysis in program evaluation* (Vol. 107). New York, NY: Jossey-Bass.

Theory of change approaches also are relevant for examining systems initiatives focused on connections. The theory of change may, for example, identify what incentives must exist for system components to connect. Then, evaluators can determine whether the theory of change is “working” by collecting data on whether the identified incentives are in place, whether connections are occurring, and whether the two can be plausibly linked.

If the evaluation’s focus is on assessing the second question of whether systems that feature connections produce better results for beneficiaries, then experimental or quasi-experimental designs may be constructed that compare conditions where connections are made to situations where they are not. Evaluations that examine connections and their impacts for beneficiaries are particularly important for systems initiatives that include a focus on developmental or subsystem-to-subsystem transitions (e.g., from preschool to kindergarten; from school to adult life for individuals with disabilities; or community re-entry strategies for ex-offenders).

Examples

The SPARK Initiative evaluation, conducted by Walter R. McDonald & Associates, is an example of an evaluation that focuses on connections. It is a cluster evaluation, meaning it involves an overall evaluator to lead the evaluation and do an “initiative level” assessment across participating SPARK sites, and project-level evaluators who work at the individual site level.

The evaluation’s design is based on a theory of change that features effective partnerships and leaders working to *align* community systems and increase the readiness of children, schools, and communities. The concept of alignment here goes beyond the notion of basic social relationships. Alignment focuses on the school’s relationship to the community; communication between early care and education programs and schools; and consistency between the early care and education curriculum and school curriculum and standards. Alignment strategies include improving early care and education, implementing developmental assessments to improve early care and education support, improving transition practices, strengthening public awareness and public will, and changing school culture. The levers of change for creating alignment are 1) partnerships within the SPARK sites and 2) an intentional leadership development effort involving key partners who become agents of local, state, and national change.

The SPARK theory of change identifies three main outcomes—ready kids, ready schools, and ready communities. Quantitative and qualitative methods assess these outcomes and the theory of change components that contribute to them. Methods include kindergarten readiness assessments, site visits (focus groups, participant observation, and key informant interviews), grantee data collection (surveys and quarterly calls), and content analysis of key documents.³⁷



³⁷ Berkley, T. (2005). SPARKing Innovation. *The Evaluation Exchange*, 11(1), 16.

4 Evaluating Infrastructure

Systems initiatives focused on infrastructure attempt to ensure the system has the supports it needs to function effectively and with quality. Better system supports may be needed, for example, in the areas of governance, financing, data systems, and practitioner support or workforce development.

Evaluation Questions

Evaluations of infrastructure-focused initiatives explore two main questions:

- 1) Did the initiative establish infrastructure or supports that are consistent with its objectives?
- 2) Did the infrastructure or supports achieve its objectives for effectiveness, sustainability, or quality?

Evaluation Methodologies

Evaluations that address the first question of whether the system has the supports it needs may use a theory of change approach. The theory of change would articulate the strategy for how a particular infrastructure change is expected to occur and then use different methods to test whether the theory of change is working as intended. Theories of change for initiatives focused on infrastructure may share characteristics in common with initiatives focused on context, as infrastructure changes often require policy changes.

Case studies offer another methodological option for these initiatives. Case studies use multiple methods (quantitative or qualitative) to examine how an infrastructure initiative played out and the factors that contributed to its success or lack thereof. For example, the Success Case Method is a particular type of case study that is relatively cost-effective. It is a post-hoc analysis of success (or failure) that combines systematic case study methodology with storytelling, and reports results that stakeholders can easily understand and believe.³⁸

Evaluations focused on the second question of whether supports are effective once they are in place examine supports' benefits *for the system*. For example, an evaluation might examine whether less categorical and more flexible funding actually results in more comprehensive and connected services at the local level. Or an evaluation might examine how practitioner supports in the form of education or training benefit the system's workforce.

These evaluations also can draw on a range of methodological options. For example, a performance audit can determine how well a governance entity or state agency is functioning. Customer satisfaction surveys can determine whether particular supports are accessible and user-friendly. Program evaluations can determine whether new education and training programs or policies have an impact on the workforce.

³⁸ Brinkerhoff, R. (2003). *The success case method: Find out quickly what's working and what's not*. San Francisco, CA: Berrett-Koehler.

Examples

The North Carolina Smart Start Technical Assistance Center evaluation provides an example for this focus area. The Center provides infrastructure development assistance to many states trying to build their own early childhood development systems. Being completed in two phases, the evaluation is exploring versions of both evaluation questions described above.

The evaluation's first phase studied the effectiveness of the Center's technical assistance in helping states to meet their system building goals. In other words, it looked at whether states were successful in meeting their infrastructure goals and the role they thought technical assistance played in their achievements. The evaluation's second phase will include a qualitative study of how technical assistance accelerated states' efforts to improve positive impacts for children. This work builds on North Carolina's own success in establishing, expanding, and sustaining Smart Start as an early childhood systems initiative in the state.

5 Evaluating Scale

Systems initiatives focused on scale attempt to ensure that a comprehensive and quality system is available to an increasing (and eventually complete) number of intended beneficiaries.

Evaluation Questions

Evaluations in this area generally concentrate on two key questions:

- 1) Did the initiative enable system scale up with quality and fidelity?
- 2) Did scale up result in broad impacts for beneficiaries at a system-wide population level?

First, evaluations can explore the *scale-up process* itself, asking whether the system is being scaled up with fidelity and quality, for example. Because there are many historical examples of systems initiatives that “demonstrated, researched, and died” during scale up, or that experienced model drift and ended up offering lesser services and benefits than intended, evaluations that focus on this process and continuously feed back findings to inform midcourse corrections can be critical to the initiative's success.³⁹

Second, evaluations can *assess the system* and its impacts for beneficiaries. The main evaluation question here is whether systems are producing comprehensive impacts for the people they serve.

Evaluations of initiatives with a scale-up focus share a lot in common with the traditional program evaluation approaches described in the earlier section about evaluating initiatives focused on components. The difference is that here the system, rather than a single program, is the “intervention.” Viewed in this way, evaluations that address the first question about the scale-up process are equivalent to program monitoring or process evaluations; those that address the second question are equivalent to program impact assessments.

³⁹ Weiss, H. (1995). New approaches to evaluating systems: The Gwen R. Iding Brogden Distinguished Lecture Series (p.412). In Liberton, C., Kutash, K., & Friedman, R. (Eds.), *8th Annual Research Conference Proceedings on A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: The Research and Training Center for Children's Mental Health.

Evaluation Methodologies

Evaluating the scale-up process, particularly when the system is designed to be available to all beneficiaries within the population (e.g. community-wide or statewide), involves some statistical analysis of the size of that beneficiary population and the degree to which the system reaches and serves them. When an initiative's ultimate goal is to produce population-wide outcomes and impacts, it is essential that the system reach a sufficient number of beneficiaries to show results at that level.⁴⁰ This requires a clarified definition of beneficiaries and the system components and connections that they will use (beneficiaries may use only a part of the system). Methods such as polls and random sample surveys can then assess the degree to which the population is aware of and uses relevant system components. Such surveys also can provide valuable feedback to systems initiative planners on where outreach, engagement, and additional focus are needed (e.g., by neighborhood, socioeconomic status, ethnicity, etc.).

Evaluations focused on the second question about system impacts can productively use a results-based accountability approach. Results-based accountability became popular during the 1990s in response to growing perceptions of public sector inefficiency and waste. This approach requires institutions at the state, community, agency, program, or system level—to articulate clearly their expected results (or goals). Once results are identified in the form of outcomes, indicators, and benchmarks, they are tracked at regular intervals to determine whether results have been achieved.

With a results-based accountability approach it is still important to construct a counterfactual or basis of comparison, as broader societal factors may influence these outcomes independent of systems initiatives. Trend data from other communities or states on similar outcome and impact measures can provide such a counterfactual for assessing whether the systems initiative itself has “turned the curve” on results or whether the changes were part of a broader societal trend.

Examples

SRI International conducted the Five 5 School Readiness Initiative evaluation from May 2002 through February 2007. The evaluation used a quasi-experimental design and collected statewide quantitative data about participants, services, and child outcomes at the time of intake and again every six months. To better manage the data across hundreds of programs and services, results were recorded in an innovative web-based system called PEDS (Proposition 10 Evaluation Data System).

SRI also conducted systems change surveys with programs, superintendents, principals, and kindergarten teachers, and completed kindergarten entry profiles every two years. Profiles provided a snapshot at kindergarten entry of children's developmental competencies and of family and community supports to help children succeed in school. Profile data were collected using two instruments—a teacher-completed checklist and a family interview.

⁴⁰ Bruner, C. (1996). Where's the beef? Getting real about what comprehensive means. In Stone, R. (Ed.), *Core issues in comprehensive community-building initiatives*. Chicago: Chapin Hall Center for Children.

Finally, the evaluation incorporated case studies of select school readiness programs into its design. These case studies provided a better picture of the initiative's scale-up process and implementation. Case studies used interviews and focus groups with staff and parents, as well as program observations.

The North Carolina Smart Start Initiative evaluation is another example in the scale category. The Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill evaluated Smart Start during its first ten years while it was being scaled up statewide. The evaluation responded to questions about how the scale-up process was working, whether Smart Start improved child care quality, and whether Smart Start was good for children. Over 35 studies were conducted during those ten years.

Evaluators used a range of evaluation designs to assess Smart Start outcomes and impacts at the program and child levels. None were experimental, although many were quasi-experimental. For example, four large preschool program samples were observed in 1994, 1996, 1999, and 2001 to document quality over time and to examine the relationship between quality increases and program participation in Smart Start improvement activities. Child school readiness data were also collected.

In addition, several qualitative studies examined Smart Start scale up through the local partnership lens. These studies explored the needs of local partnerships, their decision making processes, the challenges of involving parents and the business community, and the nature of public-private partnerships. Qualitative data helped determine where technical assistance and policy and procedure improvements were needed, and to explain how results were achieved.⁴¹

Currently, Smart Start has a results-based accountability system in the form of its Performance-Based Incentive System. Local Smart Start partnerships submit data annually on key standards and measures in the areas of early care and education, family support, health, and administration. Partnerships with excellent results are rewarded, while those not meeting expectations receive more intensive technical assistance. Also, each local partnership has its own evaluation system to example progress toward state and local goals. While state outcomes are consistent across partnerships, local outcomes can vary.⁴²



⁴¹ Bryant, D., & Ponder, K. (2004). North Carolina's Smart Start Initiative: A decade of evaluation lessons. *The Evaluation Exchange*, 10(2), 7-8.

⁴² While both California and North Carolina have committed major new resources to early childhood, it is not certain that these investments have yet been enough to create a truly complete system or even a system of sufficient scope to produce community-wide results, much less results for individual system participants.

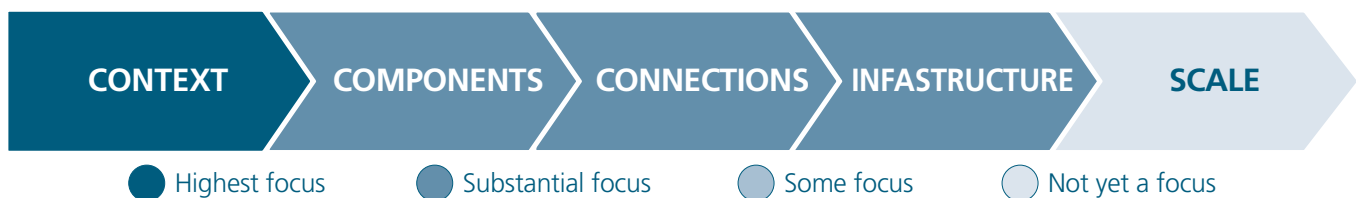
VI. Applying the Framework: The Build Initiative ● ● ●

The last two sections demonstrate how the five focus areas can act as a *framework* for defining systems initiatives, constructing their theories of change, and designing their evaluations. The five-part framework offers a way to break down the initiative's complexities into more manageable parts without losing sight of "the whole."

This section demonstrates how the framework can be applied to an actual systems initiative, with The Build Initiative serving as the example. The framework is applied to The Build Initiative in three ways; to: 1) map Build's focus areas (at an overall initiative level, across Build states), 2) define Build's relevant theory of change elements, and 3) identify Build's existing and potential evaluation options.

Mapping the areas of focus

The first step is to map The Build Initiative's focus areas and their relative emphases. This exercise sets expectations about what the initiative is trying to achieve and therefore signals where the evaluation should place more or less focus. The shaded figure below shows where the Initiative currently is focused.



To date, The Build Initiative has focused on four of the five areas—context, components, connections, and infrastructure. Of these four areas, Build states generally place the most emphasis on changing the political context to better support systems. While the Initiative's goal is to work toward scale up, no Build states are yet at the point where they are scaling up comprehensive early childhood systems.

Defining the theory of change elements

The second step is to define the theory of change elements associated with The Build Initiative's focus areas, so that the theory can further guide evaluation planning. The figure on page 28 identifies Build theory of change elements. It does not list all possible activities and outcomes; it pulls items from the theory of change menu offered earlier and identifies examples of the kinds of activities and outcomes on which many Build states focus.

Identifying evaluation options

The third step is to identify relevant evaluation options given the Initiative's theory of change. The figure on page 28 uses the evaluation design menu as a guide and bolds the elements that the national Build evaluation and local evaluation partners have focused on to date. The figure illustrates both where the evaluation is focused now and options for where it *could* focus in the future.

Note that this step necessarily requires strategic choices about what the evaluation will and will not do. For example, rarely are enough resources available to collect data on every outcome in the theory of change. Methodological choices should be based on what information the evaluation's audiences want, where the initiative currently is in its development, the evaluation's timeframe, and the evaluation resources available.

Defining Theory of Change Elements: The Build Initiative ● ● ●

● Highest focus ● Substantial focus ● Some focus ○ Not yet a focus

	CONTEXT	COMPONENTS	CONNECTIONS	INFRASTRUCTURE	SCALE
Activities (examples)	<ul style="list-style-type: none"> Developing a vision Educating policymakers Engaging the public Engaging the media Doing electronic outreach 	<ul style="list-style-type: none"> Expanding high-quality programs Developing new programs or services Expanding eligibility Increasing access and availability 	<ul style="list-style-type: none"> Creating forums for cross-sector planning Smoothing developmental transitions Connecting data systems Promoting collaboration and referrals 	<ul style="list-style-type: none"> Developing new governance proposals Improving quality (quality rating systems, standards) Improving state-local connections Developing better two-way communications Monitoring the system 	<ul style="list-style-type: none"> <i>While some Build Initiative states have focused on scaling up specific system components, scale up at a comprehensive systems level has not yet occurred.</i>
Outcomes (examples)	<ul style="list-style-type: none"> Evidence base about the system and its needs Shared early learning vision Gubernatorial and other political leadership New and mobilized early childhood advocates Changed policies (e.g., increased funding for preschool; increased child care subsidy rates) 	<ul style="list-style-type: none"> Expanded programs (e.g., universal preschool, early intervention) New programs (e.g., infant mental health services) Better access (e.g., expanded SCHIP or Medicaid eligibility) Higher-quality programs (e.g., child care) 	<ul style="list-style-type: none"> Improved transitions (e.g., 2- and 4-year college articulation) Joint training (e.g., for preschool and Kindergarten teachers) Co-located programs (e.g., "One-stop" family resource centers) Increased referrals (e.g., pediatricians to early intervention) 	<ul style="list-style-type: none"> New early childhood governance structures Quality rating systems (e.g., for child care) Assessments (e.g., Kindergarten readiness) Standards/guidelines (e.g., infant-toddler) Local infrastructure (e.g., local governance structures) 	
Impacts	<ul style="list-style-type: none"> Outcomes are not expected to show direct impacts for children and families 	<ul style="list-style-type: none"> Improved child and family outcomes in relevant domains—health, literacy, mental health, etc. 	<ul style="list-style-type: none"> Improved child and family outcomes where system components are connected 	<ul style="list-style-type: none"> Outcomes are not expected to show direct impacts for children and families 	

Identifying Evaluation Options: The Build Initiative ● ● ●

● Highest focus

● Substantial focus

● Some focus

○ Not yet a focus

	CONTEXT	COMPONENTS	CONNECTIONS	INFRASTRUCTURE	SCALE
Questions	<ol style="list-style-type: none"> Has the initiative changed the political environment through its activities? Has the initiative produced changes to investment, policy, or practice that will enable changes in components, connections, infrastructure, or scale? 	<ol style="list-style-type: none"> Did the initiative design and implement system components as intended? Did the components produce their intended impacts for beneficiaries? 	<ol style="list-style-type: none"> Did the initiative design and implement connections and linkages as intended? Did the connections and linkages produce their intended impacts? 	<ol style="list-style-type: none"> Did the initiative establish infrastructure or supports that are consistent with its objectives? Did the infrastructure or supports achieve their objectives for effectiveness, sustainability, and quality? 	<ul style="list-style-type: none"> While some Build Initiative states have focused on scaling up specific system components, scale up at a comprehensive systems level has not yet occurred.

(bolded items indicate where the Build evaluation has focused to date)

Designs/Methods	<ul style="list-style-type: none"> Theory of change evaluation Case studies Public polling Policy tracking Key informant surveys Coalition analysis Polymaker/bellwether interviews Media tracking 	<ul style="list-style-type: none"> Program evaluation methodologies (including experimental/quasi-experimental) Program monitoring Quality assessments Efficiency analyses Customer surveys 	<ul style="list-style-type: none"> Program evaluation methodologies (including experimental/quasi-experimental) System mapping Network analysis Customer surveys 	<ul style="list-style-type: none"> Theory of change evaluation Case studies Performance audits Management information systems Practitioner data collection 	
-----------------	---	---	--	---	--

This example demonstrates how the five-part framework and accompanying menus for both system initiative theories of change and evaluation designs can be used as diagnostic and planning tools. Systems efforts like The Build Initiative are complex. The framework and menus offer ways of managing that complexity so it becomes possible to see more clearly what an initiative is doing and working toward. This then aids decisions about where the evaluation should focus and how it should look.

VII. Concluding Principles for Evaluating Systems Initiatives

This paper grew from the observation that while a great deal of experience already existed on systems initiative evaluation, it was difficult to talk clearly about both this field's past and future because systems initiatives are diverse and can mean different things to different people. To address this challenge, the paper defined the different areas in which initiatives might focus and identified theory of change elements and evaluation approaches that fit with each area. The hope is that this framework will provide a clearer way of thinking and talking about systems initiatives, and therefore will facilitate learning about the field's past and decision making about its future.

While it is clear that no one "right" and generalizable way to approach evaluation has or will emerge to meet all of these challenges, this paper concludes with a set of general principles about what constitutes good evaluation practice in this area.⁴³ Some principles apply to evaluation practice generally (but still bear repeating), and others are unique to systems efforts.

1) Clarify the evaluation's audiences and intended uses for the evaluation's findings.

Systems initiatives have many stakeholders, including their funders, implementers, the programs and institutions that make up the system, and the system's intended beneficiaries. The evaluation's primary audiences (or customers) and how they intend to use it should be established upfront (e.g., for accountability purposes and determining future funding, for justifying investments made by showing results and returns on investment, or to identify lessons that can be fed back into the initiative for growth and adaptation). If there is more than one audience, all audiences' needs should be considered and balanced. While this point seems basic, it is often overlooked, resulting in evaluations that ultimately do not satisfy the audience's expectations or needs.

Once the audiences are identified, the evaluation generally should involve audience members in developing the initiative's theory of change, getting agreement about what outcomes the evaluation will assess within it, and what data will be available and when. Gaining clarity on the evaluation's purposes upfront can often help avoid misplaced expectations down the road.

⁴³ Horsch, K. (1997). Evaluating school-linked services: Considerations and best practices. *The Evaluation Exchange*, 3 (2), 2-4; Shaw, K. (1996). Challenges in evaluating comprehensive school-linked services initiatives. *Evaluation Exchange*, 1(1), 2-3.

2) Base evaluation decisions on the initiative's focus. This point cannot be stressed enough. Because systems initiatives have varied designs and areas of focus, it is not possible to discuss or decide generally how systems initiatives should be evaluated. There is no one right evaluation approach. Evaluation decisions should be based on each initiative's goals and what it is doing to reach them (as described in the theory of change). This paper was designed to facilitate these decisions, but evaluation approaches should be determined on a case-by-case basis.

3) Use theories of change to facilitate systems initiative evaluations. As stated numerous times in this paper, theories of change are now ubiquitous in systems initiative evaluations, and for good reason. Once an initiative is viewed through the clearer theory of change lens, evaluation decisions become clearer. Consequently, theories of change should continue to be a cornerstone of system initiative evaluations.

At the same time, theories of change should be constructed with rigor. They need to identify the initiative's underlying assumptions and measurable ways to test them. Until now, initiative developers (funders, leaders, collaborative groups) often have used theories of change to develop consensus on the value of embarking on a complex change process, to show how everyone's agenda fits into the overall mix, or as a framework for making decisions on first or subsequent initiative steps. While this is useful for building a shared vision and fostering collaboration, this process does not always establish the necessary groundwork for evaluation planning. Evaluators should be partners in constructing theories of change to ensure that assumptions are spelled out and that they are "testable."

4) Identify an appropriate level of methodological rigor. All evaluations need to be methodologically rigorous. Rigor applied to systems initiatives means being clear about the evaluation's outcomes, methodology, and measures. Rigor does not only mean using experimental designs that use control groups or conditions. In fact, these designs typically are not feasible with systems initiatives. This is particularly true when the goal is developing inclusive systems with open-ended participation opportunities. While the ability to make causal attributions (if the evaluation wants to make such attributions) is dependent on finding a valid counterfactual, this does not necessarily mean using randomized control groups.

The evaluation's rigor should match the evaluation question being asked. While some evaluation questions about systems initiatives, particularly those focused on components, are best answered by experimental designs, other evaluation questions call for different kinds of evaluation approaches that are more able to adapt and respond to changing initiative circumstances or strategies.

Also, methodological rigor should, to some degree, match the initiative's rigor. For example, sophisticated analyses of more "modest" systems initiatives may not be the best use of resources; they can easily overwhelm initiative efforts with reporting and documentation requirements.

5) Factor investment levels for both systems initiatives and their evaluations into evaluation decisions. Just as systems initiatives differ in focus, they also differ in funding and timeframe. Initiative funding can range from the hundreds of thousands to hundreds of millions and from a timeframe of one or two years to ten years or more. Expectations about what results initiatives will produce should be adjusted accordingly.

Similarly, initiative funding directly affects the resources available for evaluation. Standards about rigor or the “level of evidence” evaluations are expected to produce should be considered in light of their funding, as the amount of evaluation resources available greatly affects methodological choices. Generally, it is wise to conduct an evaluability assessment that assesses the proposed evaluation’s likely benefits against the cost and level of effort it creates.

6) Establish the necessary timeframe for results. Many systems initiative impacts will take numerous years to accomplish. This is particularly true for initiatives where the ultimate goal is better results for individuals such as children. These impacts may take years to show up (if they are even appropriate to measure in the first place). For example, school readiness impacts connected to early childhood development systems may require several years to show up and may only be established with real confidence when children reach third grade or higher. Evaluation plans should be realistic and expectations about results aligned with the initiative’s theory of change, its timeframe, and its scope and depth. What data to expect and when (at least roughly) should be clear to the evaluation’s audience. The fact that both those conducting systems initiatives and their evaluators will likely be pressured to “show impact” at some point before such results may be ready, underscores the importance of adhering to this principle both early on and throughout the evaluation.

In particular, systems initiatives should not be accountable for producing beneficiary impacts simply because participants agree that they would like to see such impacts occur. Measurable changes in such impacts should be expected only when a systems initiative makes investments in strategies that are powerful enough to produce them (if the theory of change’s assumptions hold).

7) Measure and *value* interim outcomes. In the context of health, education, or human services, all systems initiatives have their eyes on the ultimate prize—better impacts for the system’s intended beneficiaries. However, as stated above, these results can take many years to achieve and may involve major new investments or changes in the current system’s operations. It is important to identify outcomes that set the stage for longer-term impact, and then to avoid assigning a lesser-class status to those outcomes. While better impacts for beneficiaries should remain the ultimate goal and initiatives should be clear about how they contribute to that goal, other outcomes related to the systems initiative strategy and linked to context, components, connections, infrastructure, or scale are important measures of progress along the way.



8) Hold some systems initiatives, but not all, accountable for demonstrating beneficiary impacts. One of the most common questions about systems initiatives is whether they should be accountable for demonstrating their impacts for intended beneficiaries. This paper showed that for some initiatives, this expectation is warranted. For example, initiatives that scale up systems should be able to demonstrate such impacts. Every initiative, however, should not be held to this measurement expectation. For some initiatives—namely those focused *primarily* on context or infrastructure—beneficiary impacts are far removed from the initiative’s actual work. While the hope is that any positive outcomes these initiatives achieve will play out later for the system’s beneficiaries, and the initiative’s theory of change should clearly describe how that process will occur, it makes little sense for these evaluations to collect data that document such impacts.

9) Be clear about the initiative’s role in addressing inequities and reducing disparities. So far, most of the paper’s discussion has been at a conceptual and color-blind level, talking about beneficiaries without specifying who they might be. The term “beneficiary” was selected for broad applicability to systems initiatives, as initiatives might focus on infants, youth, families, seniors, working families, or some other societal grouping. At the same time, most systems initiatives aim to improve the well-being of beneficiaries that differ on socioeconomic status, race, ethnicity, and culture. They address the profound achievement, health, wealth, occupation, justice system involvement, and basic opportunity gaps that exist in this country.

While systems initiatives often are framed as serving “all families” or recognizing that “all children are at risk,” in reality this risk and the need for better systems responses to it are not spread evenly across populations. A fundamental goal of systems initiatives should be to redress these inequities. Some systems initiative outcomes—such as universally-available preschool, medical and family leave policies that enable parents to stay home with their infants and toddlers, and public subsidies that raise the quality of all child care arrangements—may produce benefits for children overall, but because they create “a rising tide that raises all ships,” they do not close gaps.

For both planning and evaluation purposes it is important to be explicit about initiative goals in this area, even if the systems initiative is framed to benefit all beneficiaries. Evaluations’ richest findings often are from examinations of outcomes and impacts for different subgroups in the beneficiary population. Programs benefit from this knowledge as they can see where their efforts are successful and where corrections are needed.

Also, although well beyond the scope of this paper, attention must be given to constructing evaluations that view outcomes and impacts through a multicultural as opposed to dominant culture lens. Different cultures place different values on constructs such as family, individualism, competition, nonverbal communication, and community that often are not reflected in systems that the dominant culture develops.⁴⁴ Addressing these issues goes well beyond an evaluator’s role (if systems initiatives are to address opportunity inequities and gaps, the initiatives themselves need to address ethnicity, culture, and language issues), but evaluators can help raise these issues.

⁴⁴ For a discussion of this issue with respect to “developmentally appropriate” practices within the early childhood field, see: Bruner, C. (2007) *Village building and school readiness: Closing opportunity gaps in a diverse society*. Des Moines, IA: The State Early Childhood Policy Technical Assistance Network (SECTAN); Ray, A., Bowman, B., & Robbins, J. (2006). *Preparing early childhood teachers to successfully educate all children*. Chicago, IL: The Erikson Institute.

10) Account for and examine externalities. Systems initiatives take place within and are affected by externalities—political, cultural, and other factors that are exogenous to systems initiative actions. Evaluations should take these externalities into account and factor them in when making generalizations from evaluation findings.

Especially for systems initiatives that take place in multiple locations, progress may look different in one location compared to another. Each site “is a laboratory of democracy, with its own political culture and its own set of opportunities and challenges. There clearly is no ‘one size fits all’ approach, and the sheer scope and scale of systems [work] entails many different parts (not all of which are likely to be funded or acted upon at any one time).”⁴⁵ Consequently, different sites are likely to take different systems change pathways. Results should be interpreted through lenses that include consideration of what those pathways are and the difficulty involved with systems change given, for example, the momentum for change that already exists in that site and the policy and economic climate in which change is expected.

11) Make continuous feedback and learning a priority. Regardless of design or methodological Choices, systems initiatives generally benefit from evaluations that make continuous feedback and learning a priority. Evaluators should establish adequate feedback loops to ensure timely reporting of both formative and summative findings. Evaluations that address questions raised by those who plan and implement systems initiatives can be invaluable for continuous learning and adaptation. This is particularly true with systems initiatives where a significant aspect of the work involves political will building and strategy is constantly evolving and being considered. Because systems initiatives often unfold without a predictable script, efforts to evaluate them should inform initiative strategies *as they unfold*, so that stakeholders can make good choices and identify midcourse corrections as needed. More traditional evaluation approaches in which the evaluator develops the evaluation plan and then reports back when the data are all collected and analyzed, are less appropriate here.⁴⁶

Renowned evaluator Michael Quinn Patton labels this kind of approach *developmental evaluation*, which he posits is well-suited for evolving, innovative, and transformative processes like systems initiatives.⁴⁷ Developmental evaluation features evaluators partnering

with initiative stakeholders and essentially becoming initiative team members. The evaluator’s role is to ask evaluative questions, provide data-based feedback, and generally support developmental or emerging decision making.⁴⁸

“Developmental evaluation is especially appropriate for many systems initiatives which feature high uncertainty and unpredictability, and “where the strategy is to infuse people and resources to shake up a system, increase the rate and intensity of interactions among system elements and actors, and see what happens...The very act of capturing what emerges and feeding information back into the evolving system makes this form of developmental evaluation part of the intervention.”⁴⁹



⁴⁵ Bruner, C. (2006). *Build evaluation symposium: Evaluating early learning systems building initiatives*. Unpublished manuscript.

⁴⁶ Weiss, H. (1995). New approaches to evaluating systems: The Gwen R. Iding Brogden Distinguished Lecture Series (pp. 411-416). In Liberton, C., Kutash, K., & Friedman, R. (Eds.), *8th Annual Research Conference Proceedings on A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: The Research and Training Center for Children's Mental Health.

⁴⁷ Patton, M.Q. (2006). Evaluation for the way we work. *The Nonprofit Quarterly*, 13(1), 28-33.

⁴⁸ Ibid, p.28.

⁴⁹ Patton, M.Q. (in press). Chapter 10: Conceptualizing the intervention: Alternatives for evaluating theories of change. In Patton, M.Q. (in press), *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.



**Strong Foundations For
Our Youngest Children**

www.buildinitiative.org



The Heinz Endowments, an Early Childhood Funders Collaborative member, generously supported the Build Initiative's Evaluation Symposium in April 2007 and the products produced from it, including this paper.