PROMISE NEIGHBORHOODS INSTITUTE
AT POLICYLINK

Some Considerations Pertinent to a Promise Neighborhoods National Evaluation

Authors:
Betina Jean-Louis, PhD, Harlem Children's Zone
Michael McAfee, EdD, Promise Neighborhoods Institute at PolicyLink
Lisa Cylar Miller, JD, Promise Neighborhoods Institute at PolicyLink

We would like to acknowledge the contributions of Prudence Brown, independent consultant; John Love, independent consultant; Jennifer Comey at the Office of the Deputy Mayor for Education, Washington, DC; Martha A. Rivas and Gemma North of the Youth Policy Institute; Isaac Castillo with the D.C. Promise Neighborhood Initiative; Austin Nichols at Urban Institute; Doug Imig of The Urban Child Institute and University of Memphis; and Blandina Rose of Black Family Development, Inc. These individuals, along with colleagues at the Promise Neighborhoods Institute at PolicyLink partner organizations (PolicyLink, Harlem Children's Zone, and the Center for the Study of Social Policy), and a few others who are unnamed, reviewed the document and provided important suggestions for improvement.

Please direct any questions or comments to Betina Jean-Louis: bjeanlouis@hcz.org; 212-534-0700.

Contents

Introduction

Place-based interventions—whether poorly or well executed—are among the most complex projects in social science and social policy. One that is focused on changing educational outcomes at scale and providing a full-range of supports for poor children and families is even more so. One such example is a Promise Neighborhood: a community of opportunity centered around strong schools that coordinate social, educational, health, and community services in a pipeline of supports around each child from the cradle to college to career. The U.S. Department of Education (ED) began the Promise Neighborhoods program in 2010 based on the model of the Harlem Children's Zone (HCZ), and have since provided 12 implementation grants and 44 planning grants. A national evaluation of the grantees, a constellation of diverse place-based collaboratives that are unified in their aims but differ along every other dimension,[1] will reflect and amplify the Promise Neighborhoods program's complexity and resource demands. Despite the attendant difficulties, however, such an evaluation is well worth pursuing because it provides a unique opportunity to learn about multiple aspects of place-based work, a commendable pursuit considering the number of programs and communities that now incorporate this approach (and that also struggle to address big questions relating to assessment and learning). A national evaluation would also provide unforeseen opportunities to learn about collaboration, accrual of skills and knowledge over time and across programs, relative effectiveness of different clusters of strategies and approaches to reaching the same goals, and other important variables.

This document describes some of the key issues to be considered in undertaking a comprehensive evaluation of the performance and impact of the Promise Neighborhoods grantees at large. It provides the perspective of organizations who are working at the corner of evaluation and practice—seeking to improve the outcomes for poor children and families and to understand the most effective ways of measuring results.

A national evaluation that considers Promise Neighborhoods operations, the impact on individuals who participate in the programs, and any resulting community-level changes over time, would help a number of constituents, from the ED and the White House Office of Poverty Initiatives to practitioners, participants and families, researchers, policymakers, the funding community, and the nation. Such a national evaluation would involve baseline (or early, in the absence of baseline data) measurements followed by periodic, continual investigation and feedback.

The document is organized into two major sections that cover a number of wide-ranging and specific issues pertinent to the national evaluation. **Section I** presents a number of important considerations, including the uniqueness of and need for a Promise Neighborhoods national evaluation, populations served and studied, and a timeline for expected outcomes. **Section II**

[1] In keeping with the Harlem Children's Zone's emphasis on developing approaches that address the particular strengths and needs of each community as per a core set of HCZ principles described in a white paper, "replication" of the HCZ approach inherently incorporates a great deal of differentiation across communities, an approach that is inconsistent with the usual views of program replication. Variability will necessarily exist in factors such as number and type of partner organizations involved, programs and strategies used, staffing patterns and training, definition of scale and the speed at which it will be achieved, and geography and type of community served. The white paper can be found here: http://www.hcz.org/images/stories/HCZ%20White%20Paper.pdf.

provides a review of some possibilities relating specifically to evaluation design.[2] That section is followed by a conclusion and an appendix that includes the input provided to us by members of the Promise Neighborhoods community of practice upon review of the first draft of this document.

In the course of revising this document, we elicited feedback from the practitioners and researchers involved in the Promise Neighborhoods community. The document reflects the feedback provided by these individuals. Some comments were incorporated directly into the paper and some were not. The appendix that follows the conclusion reflects the full range of feedback from these individuals, including comments that are interwoven into the paper as well as those that did not precisely align with our thinking or with the range of topics covered and were not incorporated.

The U.S. Department of Education has shown great foresight in working with the Urban Institute to lay the foundation for developing restricted-use data files (RUDFs). RUDFs are data files that are made available to researchers under guidelines designed to protect the identities of the individuals whose data are displayed therein.[3] The Urban Institute has undergone an extensive process to develop a framework for the Promise Neighborhoods RUDFs and has subsequently created a document that summarizes their data and specifications[4] and another—geared to Promise Neighborhoods grant recipients—that provides detailed and practical guidance and instructions relating to data collection, data entry, updates, and timelines.[5] The RUDFs, when completed, will include individual-level and summary data on Promise Neighborhoods participants and programs.

Making the comprehensive range of data related to Promise Neighborhoods efforts available to researchers will permit systematic evaluation. The program and the field could greatly benefit by laying the groundwork to facilitate future investigation of questions and theories that can be only partly conceived at this time. The document and guidelines produced as a result of the Urban Institute process, which included convening an impressive technical working group (receiving some input from Harlem Children's Zone and Promise Neighborhoods Institute staff members), can serve as a solid foundation for the national evaluation. Nevertheless, many factors still require consideration as decisions are made regarding how best to prepare for a national evaluation.

---

[2] Please note that the Promise Neighborhoods Institute at PolicyLink intends to devote some resources in 2014 to shaping the design of the Promise Neighborhoods national evaluation. Considered actions include commissioning the development of a plan for the national evaluation and convening a national advisory group.

[3] U.S. Department of Education, Institute of Education Sciences, National Institute for Educational Statistics. (August 2011). Restricted-Use Data Procedures Manual, https://nces.ed.gov/pubs96/96860rev.pdf.

[4] Peter A. Tatian, Christopher Hayes, and Austin Nichols, *Promise Neighborhoods Restricted-Use Data Files: Technical Specifications and Requirements* (September 2013), http://www.urban.org/publications/412909.html.

[5] Jennifer Comey, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher Hayes, Kaitlin Franks, and Reed Jordan, *Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results* (February 2013), http://www2.ed.gov/programs/promiseneighborhoods/pndataguidance.pdf and 5/21/13 at http://www.urban.org/publications/412767.html.

Section I: Important Considerations

The Unique Contributions of a Promise Neighborhoods Study

An evaluation of the Promise Neighborhoods program, in all its uniqueness and complexity, promises to shed light on unstudied or understudied but important variables relating to the success of multicomponent interventions. The variation in approach across sites married to the commonality in measured outcomes makes PN particularly compelling and potentially tremendously instructive. The program—and its evaluation—provides unparalleled opportunity to address really important and previously underconsidered social and educational science questions.

Many of the possible distinct research contributions of a PN evaluation relate to the initiative's role as a fulcrum that brings together and leverages partners, programs, and staff members with a range of knowledge and expertise. This allows, and in fact, may well mandate that the evaluation move beyond the usual approach of focusing on one particular strategy or program. Paying attention to the ties that link best practices in the pipeline (once they each have indeed been determined to be effective) really allows Promise Neighborhoods—and its evaluation—to make an enduring contribution to our understanding of how to end intergenerational poverty in long-blighted communities.

Some examples of the types of novel questions that could be asked in a Promise Neighborhoods study include:

- To what extent did the communities reach saturation with regard to participants/residents vis-à-vis particular strategies and how/does the degree of saturation affect outcomes?
- How does one create a successful place-based initiative (how does one create a pipeline and connect relevant services with real power; how does one use a place-based approach to reach and serve the most challenging individuals and families, etc.)?
- Which particular strategies or clusters of strategies did communities with better outcomes for children and families utilize most often?
- How do partners and collaborators govern themselves, hold each other accountable, and make decisions as they get their initiatives up and running?
- What staff hiring, on-boarding, training, and/or retention practices across age groups and partner agencies are correlated with successful outcomes for children and families?
- How do some of the contextual factors that impact the early implementation (like political, social, and cultural barriers to collaboration) manifest and get resolved?
- What are the infrastructure demands for a successful initiative? What are the timelines and processes for building the necessary programmatic, operational, and systems capacities to support the development of an effective Promise Neighborhood?
- How can policy guidelines relating to neighborhood-based and/or anti-poverty initiatives be informed by lessons learned from Promise Neighborhoods? Do some approaches or methodologies work better for certain segments of the population than for others (i.e., resulting in gender, linguistic, or cultural differences)? Are there any particular subpopulations that seem to be "left behind" by the initiative?
- Do benefits continue to accrue for as long as children remain in the pipeline or do gains plateau or reverse over time?

These are just some examples of the types of research questions that a multicomponent program with well-articulated logic models[6] will allow researchers to consider. The right kind of analysis can tell us a great deal about what happened at each site and why, consider collective impact with more insight into the interplay of strategies and factors than is usually possible, and give ED and the field systematic ways to compare, contrast, and learn broader lessons from multiple sites. Carefully constructed descriptive studies that explore areas that seem particularly promising could serve the dual purposes of helping to improve program functioning and generating more refined research questions and evaluation plans.

Investigating the Glue that Holds the Pipeline Together

While many community-based organizations have a great deal of experience implementing individual stand-alone programs, relatively few have run well-articulated systems of inter-connected programs. Given Promise Neighborhoods' intent to build birth through college (or high school) pipelines that incorporate the work of several collaborating organizations, the mortar that connects the individual programs and players becomes particularly important. PNI believes that this glue is a key component of success. A national evaluation would serve the program and ED well by devoting some resources to focusing on the glue that holds the pipelines together and determines the quality and effectiveness of partner collaborations. The evaluation's process study should include some consideration of the quantity and quality of deliberate efforts (including data sharing and review) that are made to promote and facilitate transitioning from one piece of the pipeline to another.

Timeline for Results

Holistic interventions such as Promise Neighborhoods operate at multiple levels of intervention and their efforts are usually intended to have impacts on individuals, families, institutions (including schools and other organizations), and communities. Promise Neighborhoods also seek to change their communities' physical and social environment but it may take years to see these changes because transformation depends on a number of factors, including: the scale at which the initiative operates, the effectiveness and focus of outreach, and the maturity of the organizations involved. Many of the problems addressed by the Promise Neighborhoods communities are so entrenched and stubborn that they require sustained effort to produce meaningful change; this, of course, accounts for the pipeline approach that allows gains down the line to build upon improvements seen earlier on.

It may be difficult to observe positive outcomes within the first few years of the initiative or during the duration of an evaluation performed too soon. Additionally, the various indicators tracked by the communities would not necessarily all show improvement at the same rate. As one example, making changes in school readiness rates at school entry may well be quicker and easier than increasing the proportion of third graders who are at or above grade level on state tests, particularly given recent and continuing changes in the level of difficulty of the tests to better align with the Common Core. Presumably, the younger children will not require as much catch up when introduced to a high-functioning educational setting than those who are in higher grades at the start of the intervention. In addition, with each additional year of the pipeline, students will be able to build upon the gains of previous years, with the knowledge and skills of each cohort of

[6] Please see the Theory of Change Design discussion in Section II of this document for some analysis of the importance of developing theory of change and logic models.

students likely to increase as ever greater proportions of the students in a given grade have accumulated more and more hours of services.

A Promise Neighborhoods national evaluation could benefit from taking explicit timelines for changes in outcomes for children, families, schools, and communities into account, considering the potential impact of multiple and progressive interventions on complex systems. The table on page 9 considers the Government Performance and Results Acts (GPRA) indicators and presents one example of how the indicators might be divided to account for differential time expectations; it is certainly possible to consider other time horizons and groupings. The really important notion here, however, is that any national evaluation that does not provide enough time for program implementation will be unlikely to show effectiveness. Realistic timelines for change will be at odds with the desire to show immediate impacts on all Promise Neighborhoods indicators.

Clearly, we recognize that a great deal of thought has been put into development of the GPRA indicators and that any investigation of Promise Neighborhoods will likely have to consider them all. However, we urge any entity that will engage in a national evaluation to pay attention to some issues that are particularly pertinent to Promise Neighborhoods as the work of developing indicator-related evaluation questions and approaches is undertaken.

To address the need for implementation time and be responsive to the desire to be able to produce early signs of programmatic success, we propose breaking the GPRA indicators into three tiers, with some indicators being expected to show more immediate gains than others. The first of the proposed time horizons begins at year 2, a deliberate choice that reflects some of the practical issues relating to disbursement of government funds and the likelihood that those funds will not be made immediately available at the beginning of the grant period.[7]

The table presents the GPRA indicators with some important adaptations. The revisions include:

- The absence of GPRA measure 11 relating to student mobility rate. While this is clearly one of the 15 indicators that ED has deemed important for Promise Neighborhoods, we could not determine what could be considered a reasonable timeline for this variable to change as a result of sustained and directed effort. Because there are so many factors external to the best efforts of any Promise Neighborhoods community that can influence mobility, holding the communities accountable for changing this in a particular timeframe seems questionable.
- Breaking up two of the GPRA measures with longer horizons into shorter- and longer-term outcomes to allow for early investigation while being respectful of the need for sufficient time to effect change. GPRA 4 appears three times and GPRA 13 appears twice, with slightly different age groupings.
- Inclusion of an additional variation of GPRA 7 that removes the phrase “matriculating without need for remediation.” Given the recent and continuing focus on the high proportions of students who enter college with insufficient knowledge and skills—requiring passing subject-based assessments, taking non-credit bearing remedial courses, or both—we understand why a high value has been placed on college readiness. Assessment of readiness helps to gauge the health of secondary education and can facilitate some very important discussions in Promise Neighborhoods communities about

[7] This timeline assumes that additional Promise Neighborhoods grantees will be selected. The current grantees are all likely to be further along when and if a national evaluation begins.

the strength of local educational systems. Gauging a community's ability to get students to college ready for the attendant challenges and able to graduate is an important goal that would be well worth measuring and achieving.

We do believe that it would be worthwhile to also consider college graduation rate decoupled from the state of readiness at entry. Community organizations that struggle to get underprepared students through college should not have that work held against them, and we are concerned about an indicator that disincentivizes promoting college attendance and graduation for all. Assessing achievement of an adapted version of GPRA 7 (GPRA 7B) that considers the degree of success in helping students to obtain degrees regardless of their readiness upon college entry would signal an acknowledgment of the value of that work.

Expected Timeframe for Change	**Government Performance and Results Act Indicators (GPRAs)**
2-5 years	- GPRA 1. # and % children who have a medical home - GPRA 2. # and % of 3-year-olds and K students who show age-appropriate functioning at beginning of program or school - GPRA 3. # and % participating in center-based or formal home-based early childhood program - GPRA 5. School attendance rates of 6th, 7th, 8th and 9th graders - GPRA 12. # and % of families of birth to K students who read to them - GPRA 13. # and % of families of K students who encourage their child to read books outside of school
6-8 Years	- GPRA 4. # and % students at or above grade level in state math and ELA assessments (3rd through 5th) - GPRA 8. # and % of children who participate in 60 minutes of physical activity daily - GPRA 9. # and % of children who consume five or more servings of fruits and vegetables daily - GPRA 10. # and % who feel safe at school and traveling to and from school - GPRA 13. # and % of families of K-5th grade children who encourage their child to read books outside of school - GPRA 15. # and % of students with school and home access to broadband internet and a connected computer
9-15 years	- GPRA 4. # and % students at or above grade level in state math and ELA assessments (3rd through 8th and at least once in high school) - GPRA 6. HS graduation rate - GPRA 7. # and % enrolled in two-year or four-year college, and receive a degree certificate, matriculating without need for remediation - GPRA 7b. # and % enrolled in two-year or four-year college, and

Expected Timeframe for Change	Government Performance and Results Act Indicators (GPRAs)
	receive a degree certificate - GPRA 13. # and % of families of K-8th grade children who encourage their child to read books outside of school - GPRA 14. # and % of parents of 9-12th graders who report talking with their child about the importance of college and career

Work and Paperwork Reduction

With the support of the Promise Neighborhoods Institute at PolicyLink (PNI), a mandate from ED, and a keen desire to do well by their participants, communities are already hard at work on their local evaluations. These site-specific efforts will provide ongoing feedback to program staff and partners that can be used to improve effectiveness and responsiveness. Information from the local evaluations must feed the national evaluation. ED can maximize resources by co-designing a dual-focus evaluation that is cohesive and integrated with the work already underway at the local level. It would be unfortunate if the national evaluation imposed additional and inconsistent demands on the communities with regard to data collection (for example, requiring that they use a different instrument than the one that had been in use by multiple communities over the years). PNI recommends that the national evaluation should build upon the work of the local evaluation to as large an extent as possible, a position that is consistent with ED's signals thus far.

Focus: Students, Participants, *and* Residents

PNI believes that Promise Neighborhoods efforts must serve a large enough proportion of underperforming children in poor communities (at least 65 percent of poor children) to create a "tipping point" and change the nature of the community,[8] redefining what success means for all.[9] The only way that the collaborations can move beyond the laudable effort of saving small numbers of individual children within a broken community to the more transformative goal of saving thousands of children in a reformed one is by reaching scale while focusing efforts on their neediest children.

As a given Promise Neighborhood community's scale increases with time—and provided that the site includes effective programs—so too does the likelihood that the neighborhood as a whole will be changed for the better. As proportionally more of the residents are served, individual-level changes should begin to aggregate up and be reflected more broadly in the neighborhood at large.

In many communities, because school choice results in a discrepancy between children who live in the designated area and those who attend the schools located therein, dedicated efforts will have

[8] Harlem Children's Zone. (n.d.). Whatever It takes: A White Paper on the Harlem Children's Zone. Accessed 5/13/13, http://www.hcz.org/images/stories/HCZ%20White%20Paper.pdf.

[9] At the end of a successful neighborhood-change effort, success for individuals and families should no longer be equivalent to moving to a "better" neighborhood.

to be made to connect with resident children who are students elsewhere. At HCZ, because there are only a few traditional middle and high schools in the zone, specific efforts must be made to reach out to the hundreds of local adolescents who do not attend the two agency-run charter schools. HCZ's multiple youth development programs provide that engagement. Less than 15 percent of the youth who receive agency middle school and high school programming are students in the HCZ Promise Academy charter schools (and thus receive supports during the school day) with most participants scattered in dozens of schools across New York City. Nevertheless, those children are engaged in high-quality academic programming in HCZ after-school programs and staff members are held to account for their progress through middle and high school and their applications and entry into college. It would not be possible to effect the desired community change if these youth, who play such a large role in determining the feel and functioning of the community, remain unserved.

Focusing on turning a critical mass of the students and young residents in the neighborhood into program participants is important for community change. In relation to the research, recognizing and mastering the challenges related to collecting data for all of these groups is important for properly documenting that community change. The guidance for the restricted use data file addresses some of the complications attendant with the file: while the GPRA indicators measure place and may include some children who may not directly participate in the programs offered, the restricted-use data file is currently set to include information for participants only. Until this is satisfactorily addressed, the best approach is for communities to collect information at all levels and make sure that the data sources are very clear about which particular group is included in each file.

Promise Neighborhoods' laudable intention to make use of pre-existing data collection processes to address some of the community-level variables further complicates the issue. Few communities are fortunate enough to serve locally recognized distinct neighborhoods that are also consistent with the boundaries by which externally collected data are delineated (e.g., zip codes, school boards, community boards, or census tracts). As a result, the defined area of focus of the Promise Neighborhoods grantee is unlikely to align with the footprint of the data that may be made publicly available. It is important to ensure that the national evaluation is able to include measures relating to participant outcomes, data pertaining to community-level impact, information about the performance of the students in the designated schools, *and* estimates of the amount of overlap in the collected data.

Dealing with Mobility

Mobility holds tremendous possibility of clouding or complicating any national evaluation of Promise Neighborhoods, particularly because mobility rates are higher among low-income households, renters, and younger families, the core audiences for the programs. A recent investigation of outcomes in the Making Connections communities found that changes in poverty rates were primarily attributable to mobility, not to changes in movers' circumstances.[10] Given that many neighborhoods experience cycles in which blocks of more affluent or poorer families transition in or out and that these changes may impact neighborhood profiles, any national evaluation will have to consider carefully how to measure and control for mobility.

[10] Claudia J. Coulton, Brett Theodos, and Margery Austin Turner, "Family Mobility and Neighborhood Change. New Evidence and Implications for Community Initiatives" (2009), http://www.urban.org/url.cfm?ID=411973.

Guidelines for the Restricted-Use Data Files

As previously discussed, ED engaged the Urban Institute to work on creating guidelines for restricted-use data files that could be made available to researchers in the future. Given the focus on implementation and funding constraints, the creation of such a file is laudable and allows for the possibility of invaluable knowledge gains as creative and disciplined researchers obtain access to such critical data. The fact remains, however, that the first two sets of implementation communities applied for and received the grants without knowing that they would be required to include their data in a data file that would be available to any and all potential researchers (as compared to a carefully vetted individual or group selected by the department). Community-based organizations understandably have many concerns relating to research[11] and can be leery of efforts that strip them of the power to tell their own story and the authority attendant with being the acknowledged expert on their work.

Administering such a data file brings a great deal of responsibility and more thought will have to be given to how best to do so. The approach of the Consortium on Chicago School Research provides one example of a comprehensive—but time-intensive—model for fostering research that supports real educational change.[12] The Consortium's steering committee provides a sounding board for researchers who seek to engage in investigation that involves Chicago public school data. The steering committee reviews research plans, shares feedback on the interpretation of results, and provides guidance on disseminating findings to practitioners, academics, policymakers, and the public at large. The committee has forged a strong relationship with Chicago public schools by creating a "no-surprises" policy that ensures that the district is not blindsided by any reports deriving from investigation of the database. While negative findings are not suppressed, keeping the district apprised of the progress of the work throughout the research process gives stakeholders the opportunity to process findings and to prepare for the potential public onslaught associated with questionable findings.

Again, the Consortium model is just one approach to consider; the point of this discussion is not to offer this particular approach as the only viable one but to highlight the need to give some additional and focused consideration to the matter of the administration of the restricted-use data file. PNI advises that ED develop carefully considered guidelines with regards to the use of this unique database that will have been developed under an unusual set of circumstances. Instead of simply making the planned restricted-use data file available to any individual who completes an application, there could be a group of invested stakeholders that include researchers, practitioners, and policymakers, who support and guide the research efforts of investigators. A group that gives serious concern to the process for making use of any publicly available data would add tremendous value. Whether or not this particular approach proves practical or appealing will require additional consideration.

[11] Harlem Children's Zone (December 2012) "Successful Research Collaborations: Rules of Engagement for Community Based Organizations," http://www.hcz.org/images/Rules_of_Engagement_paper.pdf.

[12] The Consortium on Chicago School Research: A New Model for the Role of Research in Supporting Urban School Reform" (2009), http://ccsr.uchicago.edu/sites/default/files/publications/CCSR%20Model%20Report-final.pdf.

Section II: Review of Possible Evaluation Designs

Determining the most appropriate evaluation design for an inclusive place-based initiative that incorporates community building and is intended to deliver educational, familial, and community outcomes is a highly complicated endeavor and a daunting task. However, the potential rewards of producing a high-quality and well-regarded analysis or series of analyses are large enough in magnitude to make it all worthwhile. While many communities and programs have focused their attentions on place-based work, there is still much that needs to be determined in deciphering why, how, when, with and for whom, and under what circumstances they can be made to work.

While this section presents some important discussion points relating to the national evaluation, it should be noted that this represents just a few preliminary thoughts relating to the national evaluation plan. PNI intends to devote substantial resources in 2014 to supporting the development of a proposed national evaluation plan for Promise Neighborhoods implementation grantees. In due course, we will be able to present more detailed and comprehensive thoughts relating to the desired evaluation.

In order to develop a true understanding of Promise Neighborhoods' effectiveness, it is essential to collect process *and* outcome data over the course of the intervention. Implementation data addressing the fidelity of each component to intended activities and approaches will help to frame results. For example, an intervention that results in poor outcomes could do so because (1) the program was well implemented but is not the right fit to produce the desired changes for the population under consideration or (2) the program was not implemented as required and the resulting poor outcomes are representative of improper use of an otherwise effective intervention. In the absence of process data, it would not be possible to delve deep enough to determine which of these scenarios prevails.

The contextualization made possible by a process study is particularly important for a multi-site evaluation, where one might conceivably see different results in different communities despite the use of similar strategies or approaches. Considering process is also advisable in interventions that combine multiple programs, particularly since they may all operate at varying levels of effectiveness, adding yet another wrinkle to any evaluation. A recent document describing recommendations for an evaluation of Arizona's First Things First's initiatives and programs placed a great deal of emphasis on the need to collect implementation data, recommending that any evaluation begin with determining whether each of the programs shows implementation fidelity.[13]

Given the focus on accountability and results by ED and the administration, no appeal needs to occur for inclusion of an outcome study. The Urban Institute work includes a great deal of specific suggestions regarding how best to set up data structures to facilitate investigations into process and results, so they will not be belabored here.

[13] The similarities between Arizona's multipronged approach to achieving school readiness and Promise Neighborhoods' more extended cognitive goals makes this paper worthy of review. The document is entitled "Report of the Early Childhood Research and Evaluation National Advisory Panel" and can be found here: http://www.azftf.gov/WhoWeAre/Board/Documents/National_Advisory_Panel_Report.pdf.

Robin E. Smith of the Urban Institute has written a useful brief on evaluating Choice and Promise Neighborhoods that clearly and succinctly presents some potential evaluation designs for such complicated and large interventions.[14] She suggests five possible evaluation strategies:

- Performance management
- Process study
- Experimental design
- Quasi-experimental design
- Theory of change

We present some of the pros and cons associated with each design.

Performance Management

Performance management is most akin to the work that is already being done with the local level evaluations. Here, "program managers assemble and review selected performance indicators on a recurring and frequent basis … and use these measures to adjust resource flows and make midcourse corrections" (p. 4). Because the data are reviewed and used by managers and staff, this type of evaluation tends to get greater buy-in from practitioners. Performance management is very useful in addressing the shorter-term and more limited questions relating to implementation and outcomes but is not instructive with regard to big picture considerations of effectiveness and success over time. Some of the same data used in performance management can be used in support of the national evaluation.

Process Study

While stand-alone process studies have their place, a national evaluation of Promise Neighborhoods—a program with a relentless focus on outcomes—is not one of them. Nevertheless, as discussed previously, a process study is an important component of a comprehensive evaluation of the initiative. Process studies are particularly valuable when accompanied by an investigation of outcomes in such a way that outcomes can be linked to process variables. This might be a particularly useful place to select a few case studies, given how resource-intensive process studies can be. Even though mixed design inclusive of process and outcome components would not demonstrate causation, data that illustrate clear linkages between implementation event and activities and important outcomes could prove extremely valuable to communities, policymakers, and the field.

Experimental Design

Considered the "gold standard," an experimental design in which participants and non-participants are randomly assigned is complicated enough for basic, discrete programs; for a multicomponent, extended, community-change intervention such as Promise Neighborhoods, it is impractical. Beloved by researchers because it is really the only design that permits definitive statements about causation, an experimental evaluation of Promise Neighborhoods would be extremely expensive and inconsistent with the program's intent to reach as many at-risk children in the neighborhoods as possible.[15]

---

[14] Robin E. Smith. "How to Evaluate Choice and Promise Neighborhoods," *Perspectives Brief 19*, (March 2011), http://www.urban.org/publications/412317.html.

[15] It should be noted that some individuals do not consider random assignment to be a viable approach for a place-based initiative under any circumstances. As one example, please see Austin Nichols's document *Evaluation of*

Experimental evaluations are so resource-intensive because individuals and communities that are not chosen as participants (or discontinue participation in program) generally require a great deal of incentivizing to ensure their continuing engagement. Even if one chooses "place" as the unit of randomization, the 12 communities that have already been given Promise Neighborhoods implementation grants would not be able to participate in the evaluation. Future attempts to randomize by community would run the risk of funding some communities that are not as well-prepared to implement as others, a real issue with limited resources. The resources requirement to ensure participation of non-participants is made exponentially larger by random assignment on the community level. Finally, it is unlikely that enough pairs could be assigned to have sufficient analytic power to make any determinations.

Given the additional respect and attention that is accorded to the results of random assignment studies, it is understandable that many ED administrators are working to determine how such an evaluation might be made to work within the context of Promise Neighborhoods. We are sympathetic to that view. If there is a random assignment component, we would advise against a reductionist approach that focuses on one particular program, making the evaluation of this rich, multicomponent and multiplayer program into an investigation of the efficacy of one particular intervention within the pipeline. As previously discussed, the beauty of Promise Neighborhoods lies in its holistic approach; an evaluation that doesn't tap into that approach would be a wasted opportunity. One possibility could be a multiyear and multiprogram study that randomly assigns children at one point of entry (for example, the beginning of the early childhood program) and then follows the intervention and control groups (as defined by participation in that initial program) as they then make their way through subsequent programs in the pipeline for some period of time.

Quasi-Experimental Design

Quasi-experimental designs are similar to experimental designs in that they both have a non-participating group of individuals to whom those in the intervention can be compared. The comparison group in this approach is not randomly assigned, however. Quasi-experimental designs seek to find non-participants that are similar to participants along a number of important variables (such as race, ethnicity, parental education level, amount of time living in the community). Unfortunately, when choosing individuals from the same community, the very fact that the non-participants did not present themselves as program hopefuls at one point (or presented themselves at a later date) suggests important underlying differences in motivation and other important factors between the groups, impugning the reliability of the study's findings. Because of these potential differences, causal inferences cannot be made. Quasi-experimental designs have the detriment of being similar in costs to experimental designs (since there are non-participating entities that must be tracked) while having much less authority and respect.

As with the experimental design, one could choose to find matching communities instead of matching on the individual level but then the same issue of not being able to assign enough pairs for sufficient analytic power reoccurs. In addition, it is extremely difficult (if not impossible) to find sister communities that are similar enough to satisfy all potential critics. Years ago, HCZ

Community-Wide Interventions (July 2013), accessed at http://www.urban.org/UploadedPDF/412855-Evaluation-of-Community-Wide-Interventions.pdf.

considered the possibility of a quasi-experimental study that would compare individual- and neighborhood-level changes over time to those in a similar community. Review by several consultants suggested Bedford Stuyvesant in Brooklyn, New York, as the best match. Bed-Stuy was similar to Central Harlem in racial make-up, in school outcomes, and in proportions of poor children. However, subsequent discussion by an expert panel pointed out multiple ways that the communities differed (the proportions and home countries of the immigrant populations, infant mortality rates, the strength of the community-based organizations and organizational infrastructure in the two communities, for example) and made it very clear that study results of any kind—in whichever direction—would be questioned due to these community differences. The panel suggested that believers would continue to believe and doubters would continue to doubt and that the funds could be best spent otherwise. As a result, HCZ decided not to undertake a quasi-experimental evaluation.

Theory of Change

The last approach proposed by Smith of the Urban Institute is a theory of change (TOC) design. This design "first focus[es] on the desired outcomes, then develop[s] logic models of what is needed to achieve those outcomes. Evaluators interpret program performance data within the context of the logic model, comparing actual experience to expectations" (p. 7). While PNI does not believe TOC to be practical as a stand-alone design, the TOC approach is an attractive component of a broader evaluation because it grounds the work in concrete discussions about inputs, activities, outputs, short-term goals, and long-term goals, which would ideally occur before any work actually began. Creating TOC models for the national evaluation and the individual PNI sites can be an important element in ensuring that appropriate process and outcome data are collected.

HCZ's experience working with Mathematica Policy Research to ground its longitudinal study and of the Healthy Harlem anti-obesity initiative evaluations in TOC models exemplifies that the work can be time intensive, requiring an upfront commitment as well as periodic investments. A great deal of time is required by program staff and evaluators to build valid, logical, and grounded TOC models and to keep them current as day-to-day realities lead to changes in program approach. The local evaluation team could play an invaluable role in monitoring and disclosing modifications that would then be used in making mid-course corrections of the national evaluation.

Recommendation: The Promise Neighborhoods Program Requires Design Fusion

In the end, all of the proposed individual designs have their drawbacks and it is clear that the best approach will likely involve a fusion of multiple designs. Performance management data from individual communities could point to promising practices that merit additional investigation. HCZ's experience with the Healthy Harlem Get Fit adolescent anti-obesity program is instructive in this regard. After several years of data collection revealed that the program activities were correlated with weight loss in overweight and obese children, the organization retained Mathematica Policy Research to undertake an investigation of the program using random assignment (delaying services to some overweight or obese children for a year). This is the first time that the agency has voluntarily chosen to engage in random assignment[16] and it serves as a good example of an instance where more limited data collection and review led to a gold standard evaluation.

[16] The Dobbie and Fryer studies of HCZ Promise Academy charter schools took advantage of New York State's requirement for a charter school lottery.

Conclusion

Given the complexity of the program itself and the attendant limitations of any study associated with it, ED will likely have to manage expectations relating to the potential scope and heft of a national evaluation. Evaluation is extremely expensive. The cost for evaluating Healthy Harlem, HCZ's cluster of anti-obesity prevention and treatment programs, exceeds $1M per year. This figure only includes Mathematica's costs and does not reflect substantial in-kind support from HCZ evaluation and program staff relating to staff training, observations, data entry, and staff management and support. Study costs are also alleviated by the fact that all participants are in the Harlem community, data can be collected during regularly scheduled program activities, and pre-established and ongoing relationships with individuals reduce the need for participation incentives. A Promise Neighborhoods national evaluation may likely benefit from some of these cost-reducing factors but likely not all.

Nevertheless, as has been repeatedly stated throughout this document, the national evaluation is well worth pursuing. The potential of the Promise Neighborhoods program itself for changing and refining the ways that poor children and families are supported throughout our nation can only be fully realized and extended by a national evaluation that maximizes knowledge development. PNI is pleased to have an opportunity to share its thoughts in relation to the evaluation and hope that it helps to inform ongoing discussions.

Appendix

Additional Comments and Feedback
This appendix reflects the full range of feedback received on this guidance document—comments incorporated into the paper as well as some that did not precisely align with our thinking or with the range of topics covered and were not incorporated. Comments are ordered so as to be consistent with the paper's outline, and were not necessarily received in this order.

Introduction

"Please explain the Restricted Use Data Files. I am not sure of the audience for this, and if they will know what the restricted-use data file is meant to do."

"We believe some caution is warranted with regard to the RUDFs mentioned early in the piece. With the requirement that only data on consented children/families be shared, we have concerns about whether these data will be fully representative of the services being offered to and taken up by PN families. Perhaps you have information about the consent process that leads to a different outlook. We fully agree with your concerns about surprising PNs with findings from the RUDF analyses. It is our policy to work with our partners prior to findings release to avoid surprises and also to make sure that our story resonates with—and is informed by—on-the-ground reality. "

Section I

Community Cultural Match and participatory research are other possible options in relation to the evaluation design."

The Unique Contributions of a Promise Neighborhoods Study

"Can we not *use the word "experiment" here?*[17] *You don't want to seem at all like you are advocating for an experimental design or RCT—and using this word might get people thinking in this way."*

"There are a couple of issues that don't seem very well addressed in most of the PN material. Chief among these is the multi-intervention, multisite, multiyear nature of the project, and the call to both replicate the HCZ approach and also to improvise/innovate as needed. This poses an interesting set of puzzles. We've had an interesting time, for example, trying to conduct a neighborhood survey this fall (in a small, rural, historically racially divided community). All sorts of

[17] "The variation in approach across sites married to the commonality in measured outcomes makes PN a naturally occurring experiment that can be tremendously instructive. The program—and its evaluation—provides unparalleled potential to address really important and previously underconsidered social and educational science questions" (page 6).

ideas about statistical randomness have come up against issues of who will open the door for whom."

Some additional questions pertain to:

- Whether communities are using culturally relevant methodologies in choosing and applying their interventions, and
- Additional research and policy imperatives.

Timeline for Results

"We appreciate your focus on implementation issues and questions that are so important for understanding initiative success. Much of our work in the first few years is implementation research. We wonder whether a national evaluation that does not start at the beginning of the implementation process will be able to capture that information. The potential for information to be lost is tremendous if the evaluation begins substantially after initial implementation. We also share your concerns about timeline with regard to expectations about movement on the GPRA indicators."

"I'm very interested in the proposed timeframes for change that you suggest in the paper. Where GPRAs involve more systemic cultural change, I agree that meaningful improvements will take some time to realize.[18] This is also complicated by shifting measurements with moves toward the common core and shifting state accountability tests."

[With regard to the discussion about mobility on page 8 and advocating removal of the related indicator]:

"Is this not the overall vision = community stability?"

"We should not remove the indicator relating to mobility despite the likely difficulty of attaining the goal because retaining it will help to ensure that the communities keep it as an area of focus and coalesce around strategies to address it."

"I would suggest changing the timelines associated with the expected timeframe for change in the following ways:

- *Move GPRAs 8, 9, and 13 (K-5th grades) from a six- to eight-year timeframe to a two- to five-year horizon.[19]*

[18] Several PN communities are dealing with political realities which further complicate systemic change such as the school district being in conservatorship/receivership, multiple changes in leadership at the superintendent and board level, etc.

[19] GPRA 8 = # and % of children who participate in 60 minutes of physical activity daily; GPRA 9 = # and % of children who consume five or more servings of fruits and vegetables daily; GRPA 13 = # and % of families of K-5th grade children who encourage their child to read books outside of school.

- *Move GPRAs 6, 13 (K-8th grades), and 14 from a nine- to 15-year timeframe to a four- to six-year horizon.[20]"*

"I'm not fully clear on the reasoning behind changing GPRA 7 to remove the phrase "without need for remediation." This may be less of an issue for the HCZ, but our mean ACT score out of high school is a 15. Our kids need serious remediation or seriously stronger prep to succeed in college. We feel a need to get the community on board for this point. On GPRA 7, here's another way of putting it: about half our graduates each year attempt college, but almost none of those kids are "college ready." There is a power in data to have honest conversations about what a ready child, ready family, and ready school looks like."

Focus: Students, Participants *and* Residents

"We're completely on board with your call for changing the culture in low-income communities to change the nature of the community. The call to improve neighborhoods through Promise Neighborhood initiatives also requires that there are meaningful job and life opportunities there. Again, these conversations need to come into alignment with the discussion of strengthening home and school experiences for children."

Section II

Process Study

"This would be a good place to consider case study analysis."

Experimental Design

In response to the statement, "One possibility could be a multiyear and multiprogram study that randomly assigns children at one point of entry (for example, the beginning of the early childhood program) and then follows the intervention and control groups (as defined by participation in that initial program) as they then make their way through subsequent options in the pipeline for some period of time":

"Another approach could be modeled on what was done with Promotores. Treatment group got a Promotor (in PN case, assigned to the pipeline) and the control got programming as usual (in PN case, no pipeline—or no help navigating the pipeline)."

"I would argue that such a design is inappropriate—if a comprehensive intervention is designed to work across an entire geographical area whether or not people are direct recipients of services, the spillovers mean there is no real control group."

Promise Neighborhoods Program Requires Design Fusion

[20] GPRA 6 = HS graduation rate; GPRA 13 = # and % of families of K-8th grade children who encourage their child to read books outside of school; GPRA 14 = # and % of parents of 9-12th graders who report talking with their child about the importance of college and career.

"I think this is the critical part of the paper, and it needs to be fleshed out some more. Would you say that any national PN needs to include all 5 of the approaches? And if so, to what extent and in what order? I can certainly see the need to do Theory of Change, Process, and Performance Management early on for each PN. But what then? Who decides if they should pursue a Quasi-Experimental or RCT, and at what point? I guess the big thing I am looking for here is a more detailed description of what the 'fusion' would look like. Even if the example you give, there could be some more detail about how the different types of evaluation complemented each other.[21]"

General Comments on the Overall Document

"I do not have any major feedback. Just hopes that the guidance document for the national evaluation will include pointers on 1) how to approach active consent with program participants since it will have implications on our analyses; 2) partner agency data capacity—PN demands certain levels of program and participant info but partners may not have the resources to collect and track that info; 3) Results Based Accountability integration and utilization among PN staff and partner agencies, how this will be integrated in the national evaluation."

"I really like the document as a whole—especially the time frames you proposed for movement on the GPRA indicators. I also like the discussion about resident mobility and how that will need to be figured out. Gentrification will also play a role—we are going to get poorer residents displaced by higher cost housing. That will decrease the poverty level in the community, but is this really the way we (or ED) want to decrease poverty?"

"My gut reaction was to switch/flop sections 1 and 2. Put 2 first then 1. But I can see it working either way. Rationale for putting 2 first: to quickly address any ideas of doing a more 'traditional/RCT' evaluation of the Promise Neighborhoods. But section 1 provides a lot of important context. Just think about it."

"We appreciate your discussion of the Healthy Harlem Get Fit program and have used that story a number of times to explain to our providers that evaluation can be costly and complex and time consuming."

[21] The Promise Neighborhoods Institute will support the development of more detailed and comprehensive proposals relating to the national evaluation.

PolicyLink

PolicyLink is a national research and action institute advancing economic and social equity by Lifting Up What Works®.

Headquarters:
1438 Webster Street
Suite 303
Oakland, CA 94612
t 510 663-2333
f 510 663-9684

Communications:
55 West 39th Street
11th Floor
New York, NY 10018
t 212 629-9570
f 212 768-2350

www.policylink.org
www.promiseneighborhoodsinstitute.org

PROMISE NEIGHBORHOODS INSTITUTE
AT POLICYLINK

The Promise Neighborhoods Institute at PolicyLink helps build and sustain Promise Neighborhoods to ensure that all children can reach their full potential.