Evaluating Programs and Impact within Promise Neighborhoods

A Guide to Planning Successful Evaluations

Peter A. Tatian and Benny Docter

December 2020

Place-based education and community change interventions face distinct challenges designing and executing high-quality evaluations. Because these efforts attempt to create population-level change by using a comprehensive continuum of cradle-to-career programming, impact evaluation methods such as randomized controlled trials may be impractical or inappropriate. Nevertheless, planning, formative, and quasi-experimental methods can be used to conduct rigorous and instructive evaluations of Promise Neighborhoods.

Recent research on how neighborhoods can affect children's educational and lifelong outcomes (Chetty and Hendren 2018) has made the need for neighborhood-level interventions to address structural barriers to educational and economic mobility clearer than ever before. But which kinds of programming, and in which combinations, could create the desired community-wide change? Finding answers to these questions has only increased in urgency given the inequitable impacts on education outcomes stemming from the COVID-19 pandemic and distance education.[1] Promise Neighborhoods, and other place-based initiatives, need guidance on how to evaluate the effectiveness and results of their efforts to create positive change for students, families, and communities.

In this brief, we present options and discuss best practices for Promise Neighborhoods conducting evaluations within their communities. We focus on evaluating individual Promise Neighborhoods, not assessing the impact of the entire Promise Neighborhoods program. Our review aims to advise practitioners and evaluators on the best ways to conduct an evaluation within the context of a place-based education and community change intervention.

We start with a brief overview of the Promise Neighborhood model and then discuss basic evaluation approaches. Evaluation covers a range of methods, each of which is designed to answer specific questions. Evaluators must choose the correct approach to match their goals. We also discuss the evaluation challenges specific to the Promise Neighborhoods model and describe helpful opportunities and resources. These challenges include responding to changing circumstances, such as the COVID-19 pandemic. Throughout the brief we include examples of evaluations from Promise Neighborhoods and other place-based initiatives.

Background

The US Department of Education's Promise Neighborhoods program issued its first round of planning grants in 2010, with the first full implementation grants awarded to five recipients in 2011. Since then, 28 Promise Neighborhoods have received federal implementation funding.[2] The Promise Neighborhoods program is designed around a results framework with 10 results that each community aims to achieve,[3] from children being ready for kindergarten to families and community members supporting learning in Promise Neighborhoods schools. The program also recognizes that each community is unique and provides flexibility for how results can be achieved.

From the very beginning, people have been working on how to evaluate the program and its components. Focus on evaluation increased during the 2017 fiscal year, however, when the notice of funding availability required new Promise Neighborhoods grant applicants to submit plans for an evaluation.[4] Even before that, however, many Promise Neighborhoods teams were working with their own evaluators to document and learn from their program's results.

For most of its history, the Promise Neighborhoods program had no national evaluator. In 2019, the Department of Education awarded a contract to Mathematica, Inc., to conduct a national evaluation, which is in progress.[5] Mathematica previously wrote a feasibility study for a Promise Neighborhoods national evaluation.[6]

Promise Neighborhoods face several challenges in designing and implementing effective evaluations, one of which is that no single evaluation approach works across the different communities. Promise Neighborhoods implementation grantees include both rural and urban communities, with wide ranges in demographic and economic conditions as well as geographic sizes and population scales. As a result, grantees' sequences of programming (also referred to as the cradle-to-career pipeline) can look very different, as each grantee tailors its approach to its communities and needs. This means that evaluation approaches must also be customized to each grantee's situation.

What Is Evaluation?

"Evaluation" refers to various activities that provide evidence about what a program or strategy did and how well it achieved its aims.[7] Evaluations are designed to answer predetermined questions and are usually carried out over a discrete time frame. They often require additional data beyond what are normally collected for regular performance tracking, and they are frequently undertaken by people other than the program staff.

Different types of evaluation answer different questions, so it is important to formulate clear research questions from the beginning and use the right approach to answer them. Not all evaluations measure impact directly or provide generalizable conclusions about whether a program was effective. Many evaluations gather valuable insights about what groups a program is working with, what recipients' experiences are, and how the program can be improved or standardized. Further, evaluation methods should be viewed as complementary: using multiple methods simultaneously (or sequentially) will yield a more complete picture of a program or strategy.

Formative Evaluation

A *formative evaluation* answers questions about how a program is designed or carried out. As defined by the US Department of Education, a formative evaluation "examines the implementation process, as well as outcomes measured throughout program implementation, in order to make decisions about midcourse adjustments, technical assistance, or professional development that may be needed, as well as to document your program's implementation so that educators in other classrooms, schools, or districts can learn from your program's evaluation" (Giancola 2014, 3).

One type of formative evaluation, a *planning study*, takes place during the design or planning phase. A planning study aims to clarify a program's or strategy's plans and to make improvements early on, and it can be used to address questions such as the following:

- What are the goals and objectives of the program or strategy?
- What population is the intervention intended to serve?
- Is the intervention appropriate for the identified goals and population?
- What is the intervention's expected impact? Is there sufficient evidence to support this prediction?
- Are the data being collected sufficient to document implementation and outcomes?
- Are the available resources (staff, facilities, equipment, funding) adequate to accomplish the goals and objectives?
- Is the implementation timeline achievable?

Formative evaluation can also be undertaken throughout program implementation, as an *implementation* or *process study*.[8] A process study can be particularly important for programs that are still developing so changes during implementation can be clearly documented. In addition, Promise Neighborhoods using programs that follow a specific model will want to assess whether what they implement is faithful to that model. A process study answers questions about the quality of program implementation, such as the following:

- What interventions were implemented?
- Were services delivered as intended? Did the program reach the appropriate people? If not, why not?
- Were enough data collected to measure implementation and outcomes accurately and in a timely manner?
- Did the program have sufficient resources (staff, facilities, equipment, funding) to accomplish its goals and objectives?
- Did staff members encounter problems in setting up or running the program? Were they able to respond to and address all challenges?

The Cuyahoga housing counseling program in Ohio illustrates how a formative evaluation can answer key questions about program implementation (box 1).

BOX 1

Formative Evaluation in Focus: Cuyahoga Housing Counseling

The Cuyahoga Partnering for Family Success Program sought to improve outcomes for housing-insecure families by offering intensive counseling that helped connect housing with a wide array of services. A formative evaluation, published in July 2017, explored how and when services were delivered, tested the systems of data collection, and recommended how to improve program alignment (Bai et al. 2017). The study found that, on average, clients met with counselors 35.7 times in the first year of the program and that they were 17 percent more likely to be enrolled in SNAP benefits after receiving programming than before programming. Although clients in counseling received more SNAP support, they received less emergency housing assistance, such as shelter stays. In addition to this promising quantitative data, the process evaluation used mixed methods, including stakeholder interviews, to determine how counseling was working for each stakeholder and what data were collected at each step of the process. This allowed the evaluators to make recommendations about standardizing the services administered and collecting data smoothly and more regularly. This successful formative evaluation was also a good foundation for a subsequent summative randomized impact study.

Summative Evaluation

The second major type of evaluation is *summative* or *impact evaluation*. While formative evaluations examine current program operations and results, summative evaluations take groups of people who have completed a given program and look retrospectively at what that program accomplished. The Department of Education says, “You can use summative evaluation results from rigorous evaluations to make final, outcome-related decisions about whether a program should be funded or whether program funding should be changed. Summative decisions include whether to continue, expand, or discontinue a program based on evaluation findings” (Giancola 2014, 3).

Summative evaluations are intended to answer one question: *Did an intervention produce the desired outcomes in the intended populations?* For example, did the intervention improve school math and reading test scores? Reduce student absenteeism? Increase the numbers of parents who read to their children?

The simplest summative evaluations do not use statistical controls or a comparison group to answer this question. They are sometimes known as *outcome evaluations*. The key distinction between outcome evaluations and formative evaluations is that outcome evaluations use data on the program's final intended outcomes and examine those outcomes for at least one cohort of participants who have completed the program.

A pre-post assessment structure can form the basis of outcome evaluations. Pre-post analysis is an approach to assessing changes in participants, in which evaluators measure an outcome once at the start of the program and once at the end (Black 2004). An alternative approach is a post-then-pre analysis, in which evaluators measure an outcome once at the end of the program and then ask participants to assess retrospectively whether they experienced any change as a result of the program (Rockwell and Kohn 1989; University of Wisconsin–Extension 2005). Post-then-pre is often used for training programs where someone's perception of how much they know about a topic may change as a result of the training. For example, participants in a financial literacy program may think they know a great deal about the topic beforehand, but after going through the program, they may become more aware of concepts they do not understand well.

While simple outcome evaluations can provide a general sense of whether a program or strategy is producing improvements, evaluations that only use data from among program participants, without a comparison group, cannot compare what did happen to what would (or might) have happened without the program. This severely limits any claims the evaluation can make about what the program or strategy accomplished on its own. For this reason, Promise Neighborhoods and similar organizations that are investing significant resources in a summative evaluation should find a way to develop valid comparison groups.

Summative evaluations focus on *impact*. "Impact" is sometimes used loosely to describe any observed change, but, most properly, impact refers to changes produced by the intervention alone and not caused by other factors. Impact most typically reflects direct outcomes—that is, persistent changes in participants' situations or behavior. But impacts can also include changes outside those seen in participants, such as impacts on other family members or the community. Impacts can also be intended (i.e., meant to be caused by the intervention) or unintended (i.e., incidental or even contrary to the intervention's original goals or objectives).

Proving true impact is challenging. An *experimental study*, also referred to as a *randomized controlled trial*, is generally considered the most reliable way to evaluate impact. But experimental studies are not appropriate in all situations. For Promise Neighborhoods, experimental studies may not be feasible if programming, such as tutoring or home visiting, cannot be randomly assigned to students or families. An alternative approach is a comparison study, where results for Promise Neighborhood students, families, or communities are compared with similar populations. But identifying appropriate comparable populations can also be difficult. Promise Neighborhoods communities often have unique circumstances, so finding similar households, schools, or neighborhoods is not always possible.

An alternative to an experimental study is a *comparison study*, also referred to as a *quasi-experimental study*. In a quasi-experimental study, changes in outcomes for program participants are compared with changes for a comparison group that resembles the participant group as much as possible. For example, a program to help public housing residents increase their savings might look at savings rates for program participants versus rates for other public housing residents who choose to not take part. Generally, statistical methods (such as regression models) are used to adjust for any differences between the two groups that might affect the outcome of interest, such as household income, employment status, and educational attainment. The New York City community schools program illustrates how a quasi-experimental design can be used to assess the impact of school-level programming on chronic absenteeism (box 2).

In addition to helping direct resources and scale successful programs, impact assessments can help practitioners better understand the strengths and weaknesses of their programs. Looking at participants' program data alone, it can be tempting to assume that a program is working best for the people who demonstrate the intended outcomes after completing the program. In reality, however, those students and families may not be the ones who can benefit most from the program. By using a comparison group, a summative evaluation can reveal where the biggest differences are. Often, practitioners can use insights about who is best served by their programs even if the overall effectiveness findings are unclear.

IS YOUR PROGRAM OR STRATEGY READY FOR AN IMPACT EVALUATION?

Impact evaluations can respond to questions no other kind of research can answer. Because of the complexity and cost involved, however, there are several key questions to review before deciding to proceed with an impact evaluation. Evaluation planners should be able to answer "yes" to each question below before beginning an impact evaluation. If they cannot, the team is likely better served by pursuing formative evaluation, such as a planning study or a process and implementation study, until the program or strategy is in a more suitable place for an impact evaluation.

1. Does the program/strategy have a formal design or model in place? Is the design or model sound?
2. Is the program/strategy being implemented as designed?
3. Is the program/strategy serving the intended population?
4. Does the program/strategy have the resources needed to succeed?
5. Can you produce data for an evaluation?
6. Will an evaluation yield a meaningful result for your program/strategy or the field?

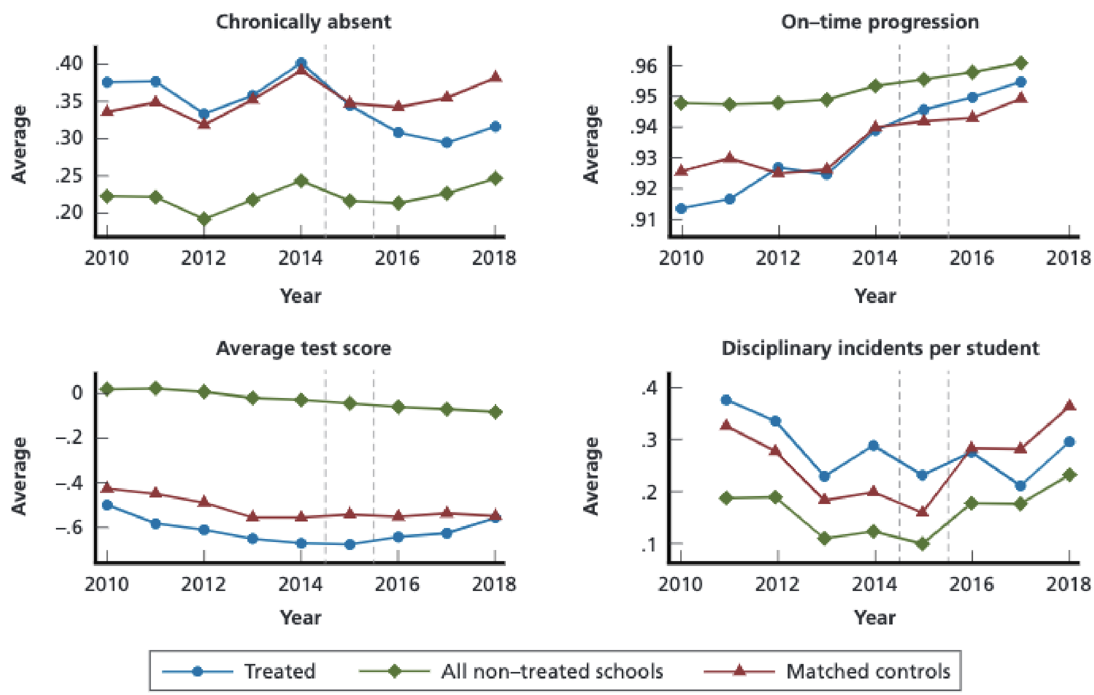
BOX 2

New York City Community Schools

Similar to the Cuyahoga housing program, the New York City community schools program started with a formative evaluation (Johnston et al. 2017). After evaluators had tested data systems and developed a more robust understanding of their programming, they used a summative evaluation to isolate meaningful school-level impacts for the community program (officially called the Attendance Improvement and Dropout Prevention Grant). Because these 118 grants were not awarded randomly, the summative evaluation used a quasi-experimental design to measure impact. Specifically, the authors tried to identify the "statistical twin" of each school using principal component analysis. As shown in figure 1, using this matched comparison group, the authors found a robust effect of the program on chronic absenteeism. Other indicators were slightly less pronounced but nonetheless showed statistical significance across a range of metrics. Many of those metrics are related, if not perfectly aligned, to the central goals of Promise Neighborhoods.

FIGURE 1

Average Outcomes of Non-Community Schools, Community Schools, and Matched Comparison Schools Over Time: Elementary and Middle Schools



Source: William R. Johnston, John Engberg, Isaac M. Opper, Lisa Sontag-Padilla, and Lea Xenakis, *Illustrating the Promise of Community Schools: An Assessment of the Impact of the New York City Community Schools Initiative* (Santa Monica, CA: RAND Corporation, 2020), 35. Reprinted with permission.

Notes: Dashed vertical lines indicate that 2014–15 is considered a transition year. The vertical scale for "chronically absent" and "on-time progression" is the proportion of students in those categories, averaged over schools. The vertical scale for "average test score" is standardized test scores, averaged over schools. The vertical scale for "disciplinary incidents per student" is the number of incidents, averaged over schools.

Evaluation Goals and Approach

It is important to identify the key questions and goals of any evaluation effort before deciding on the appropriate approach. Table 1 illustrates how different evaluation approaches are used to address particular goals. These approaches are not mutually exclusive and can be applied as needed to address key questions at appropriate points in the evolution of a program or strategy.

TABLE 1

Evaluation Goals and Approaches

Evaluation goal	Evaluation approach
Supporting education practitioners learning from one another and improving results	Formative
Confirming fidelity to program models	Formative
Establishing an evidence base to support programmatic choices	Formative or summative
Satisfying federal evaluation requirements	Formative or summative
Determining the groups for which a program has the largest impacts	Summative
Funders or policymakers deciding where to prioritize resources across programs	Summative

As the table shows, a formative evaluation often presents the most concrete path forward for practitioners just starting their evaluations. Summative evaluation may be best considered at later stages of a Promise Neighborhood's development, when programming models and data collection are more robust and mature. Nevertheless, some Promise Neighborhoods may be able to conduct summative evaluations sooner, if they are expanding upon existing programming or designing interventions with an experimental or quasi-experimental approach in mind.

When identifying evaluation goals, Promise Neighborhoods should consider using community-based methods to inform the process (box 3). Getting input from the people who are meant to be served on the evaluation questions that they would most like answered will help increase the relevance and accountability of the Promise Neighborhood to the community. Promise Neighborhoods can also apply community-based methods to any of their evaluation approaches to strengthen the relevance, interpretation, and effectiveness of their evaluations.

BOX 3

Community-Based Methods

Community-based methods incorporate the "input, participation, and reflections of the people and communities at the heart of the issues."[a] Community-based methods range from consultation, where community members have opportunities to provide input to the research process and interpretation, to community-based participation, where community members participate as full research partners, including helping formulate research questions, refine approaches, and interpret results.

Using community-based methods in evaluation not only helps address ethical issues around researching low-income communities or communities of color, but improves the quality of the research by using the lived expertise of residents in designing methodologies, collecting data, and interpreting and applying findings. The HOPE SF program in San Francisco showcases such a mixed-methods approach. HOPE SF describes itself as "the nation's first large-scale community development and reparations initiative aimed at creating vibrant, inclusive, mixed-income communities without mass displacement of the original residents." The program's baseline report outlined both the formative and summative evaluation approaches (LFA Group 2012). By bringing in community-based participatory research experts, HOPE SF's evaluation team sought to tailor research questions and methods that were responsive to the community's concerns (Jutte et al. 2011).

[a] "Community Engagement Methods at Urban," Urban Institute, accessed November 30, 2020, https://www.urban.org/research/data-methods/community-engagement-methods-urban.

Tailoring Evaluation to Promise Neighborhoods

The previous sections described general evaluation approaches, but Promise Neighborhoods face particular requirements in framing and implementing evaluations that can provide instructive results. Three of these requirements and ways to address them are discussed below.

THE PROMISE NEIGHBORHOODS CRADLE-TO-CAREER PIPELINE IS OVER 20 YEARS LONG

Promise Neighborhood solutions are intended to support children, youth, and families from birth through early, primary, secondary, and postsecondary education, to career. Fully completing this trajectory is a long-term prospect, but Promise Neighborhoods need evaluations that can provide near-term feedback. Promise Neighborhoods also aspire to produce community-wide changes for residents and students, and it can take years to see progress at that level.

To address this challenge, Promise Neighborhood evaluators need to look at both short- and long-term outcomes and create appropriate ways of assessing each within an overall results framework. Short-term outcomes, such as improvements in student attendance or grade point averages, can be tracked over a few months or years and often are preconditions for larger, longer-term outcomes. In cases where evaluators hope to gain insights about a long-term impact, such as college enrollment without remediation, these results can take longer to emerge because the layered impacts of each other part of the pipeline can take years to emerge. A solution is to use interim indicators that other researchers have established are causally connected with the desired outcome (Auspos 2012).

PROMISE NEIGHBORHOODS CONTINUALLY ADJUST PROGRAMMING TO IMPROVE RESULTS

Promise Neighborhoods are not a standard and immutable set of solutions. Rather, Promise Neighborhoods adapt their programming to the needs of their population and continually monitor progress. If new solutions are needed (such as to respond to changing circumstances, like the COVID-19 pandemic) or certain solutions are not yielding the expected results, Promise Neighborhoods adapt their approach. Therefore, Promise Neighborhoods cannot be evaluated as a fixed program model.

The adaptability of Promise Neighborhoods reinforces the importance of formative evaluation as part of an overall evaluation strategy. To be able to draw meaningful conclusions from an evaluation, the interventions must be clearly documented, including any midcourse changes. Adaptations during planning and implementation can be identified and analyzed through a formative evaluation. This information will be invaluable to plan for and undertake a later summative evaluation, either of individual components or of the overall Promise Neighborhoods pipeline.

In addition to the above challenges, Promise Neighborhoods must contend with changes taking place because of the COVID-19 pandemic. The pandemic has affected not only the delivery of services and programming, but data collection and evaluation planning. Box 4 gives some guidance on ways Promise Neighborhoods can adapt their evaluation plans in light of COVID-19.

BOX 4

Evaluation and COVID-19

Because of the COVID-19 pandemic, Promise Neighborhoods and other initiatives face new evaluation challenges around the fidelity of methods and comparability of results. The World Bank Group has some recommendations on adapting evaluation designs to the reality of the COVID-19 pandemic, including thinking through the following questions:

Should we adapt our evaluation's questions and scope?

- Will decisionmakers be willing and able to act on evaluation findings?
- Should we redirect the evaluation's focus to issues more relevant during the pandemic?
- Do we have the resources and capacity to complete a successful evaluation?

Can we improve the parts of our evaluation plan that remain feasible?

- Can we increase our reviews of current documentation and improve their usefulness?
- Do we have an opportunity to apply more multidisciplinary approaches?

Can we find ways around the parts of our evaluation plan that are infeasible?

- What can we do that does not require in-person meetings?
- Can we rely more on the expertise of community members to fill knowledge gaps?
- When reaching out to frontline workers and residents, can we use "an abundance of caution" by applying strict informed consent and ethical procedures? How can we maximize the benefit, and minimize the potential burden or harm, of our research?

Can we tap into alternative sources of evidence?

- Is there an opportunity to develop new sources of data and knowledge?

Source: Estelle Raimondo, Jos Vaessen, and Mariana Branco, "Adapting Evaluation Designs in Times of COVID-19 (Coronavirus): Four Questions to Guide Decisions," *#What Works* (blog), World Bank Group, Independent Evaluation Group, April 22, 2020, https://ieg.worldbankgroup.org/blog/adapting-evaluation-designs-times-covid-19-coronavirus-four-questions-guide-decisions.

A PROMISE NEIGHBORHOOD'S IMPACT IS MORE THAN THE SUM OF ITS PROGRAM EFFECTS

As a place-based intervention model, the Promise Neighborhoods approach relies on overlaying programming and policy changes on a focused population and area to produce spillover effects that amplify the results.[9] As a result, evaluating the individual components of the cradle-to-career pipeline will not give a full picture of a Promise Neighborhood's effectiveness.

To measure the larger impact, evaluators must find similar students, schools, or communities against which a Promise Neighborhood's results can be compared. Similar data must be collected on the Promise Neighborhood and the comparison population, on both the outcomes and the factors that can influence those outcomes. Propensity scoring is one way to identify appropriate comparison populations based on common characteristics.[10] Another approach that can be appropriate for Promise Neighborhoods is synthetic controls, where multiple observations are combined statistically to yield more robust comparisons than comparing with individual school districts or communities.[11]

Evaluation Experiences of Promise Neighborhoods

Several Promise Neighborhoods have already worked with local evaluators. These efforts illustrate how evaluation methods can be applied to the particular circumstances of Promise Neighborhoods.

Promise Heights

The University of Maryland, Baltimore–School of Social Work (UMB-SSW) team has done extensive evaluation planning. Since receiving its "Promise Heights" federal grant in 2018, the team has set up evaluation systems that will support both formative and summative evaluations. The Promise Heights team benefits from UMB-SSW being home to Maryland's Longitudinal Data System (MLDS), which tracks student-level information for every public and charter school in the state. While the Promise Heights team cannot access the full MLDS system directly, it has partnered with the UMB team that runs MLDS. By doing so, the Promise Heights team has been able to custom-frame and adjust its research questions while getting feedback from the people who know the data system best.

Separately, Promise Heights funded each of its partner schools to upgrade the district-provided student tracking system (known as i-ready). All Baltimore schools are given a basic subscription to the service, but schools in the Promise Heights footprint that agree to share their data can access a deluxe version of the software. Promise Heights uses the data to compare which in-school programs are showing the most promising signs of success. This analysis is just getting started, so Promise Heights does not yet have time-trends or statistically valid comparison groups for a summative evaluation, but the team is building toward this for the future. In the meantime, the team has extremely rich descriptive data on the academic performance and attendance of all the in-school program participants. Finally, the Promise Heights team is partnering with Urban Institute researchers to conduct an ongoing formative study focused on evaluating how community members see Promise Heights and what the program team can learn from its relationships in the community.

Chula Vista

South Bay Community Services has run the Chula Vista Promise Neighborhood in Southern California since receiving an implementation grant in 2012. Though federal grants typically last five years, Chula Vista's grant has been extended twice, giving the program more time to integrate relationships among program staff, internal data teams, and the external evaluator, San Diego Association of Governments. Chula Vista's leadership team attributes the program's success to the continuous design, planning, and implementation of programming with data and evaluation staff present.

Through this integration, the Chula Vista team has realized how important it is for data collection systems to be "lean" and for program staff to receive constant feedback on how the data they are collecting are being used. Team members meet several times a year to ensure that data systems are narrowly tailored to the three questions central to the Results-Based Accountability framework. Special care is taken to ensure that, as much as possible, data collection is limited to helping answer one of the following questions:

1. "How much?"—in other words, how many people received the programming
2. "How well?"—in other words, how many people successfully completed the programming
3. "Is anyone better off?"—in other words, what is the programming's impact

Though collecting other data could be useful, the Chula Vista team has found that focusing on these questions has increased their data's quality and reliability.

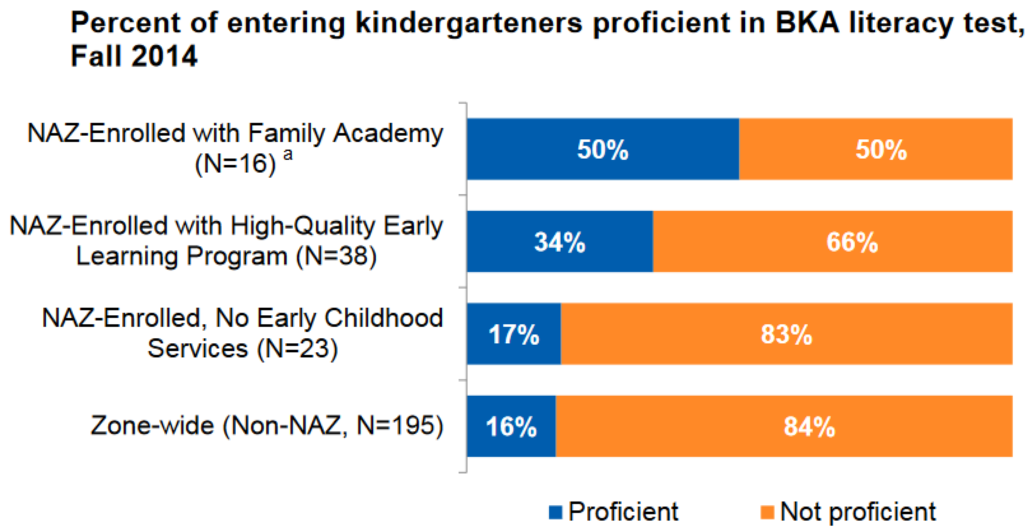
Northside Achievement Zone

The Northside Achievement Zone (NAZ) in Minneapolis received Promise Neighborhoods implementation funding from 2012 through 2017 and worked with Wilder Research as an evaluation partner. Wilder Research wrote annual reports summarizing NAZ's accomplishments, analyzing outcomes, and describing programs. Outcome data analysis included comparisons of people in specific programming with people enrolled in NAZ who did not receive services and non-enrolled peers living in the Promise Neighborhood (figure 2).

The annual reports also give useful context for the findings. For example, the discussion of figure 2 mentions that "Family Academy participation may be one of the causes of enrollment in high-quality early learning programs, or both could be caused by an underlying factor such as a parent's increased interest in involvement" (Shelton, Warren, and Gehrig 2017, 14).

FIGURE 2

Sample Results from the Northside Achievement Zone (NAZ) 2016 Annual Report



Source: Ellen Shelton, Cael Warren, and Sarah Gehrig, *NAZ 2016 Annual Report: Including Accomplishments over Five Years as a Promise Neighborhood* (Saint Paul, MN: Wilder Research, 2017), 14.

Notes: BKA is the Beginning of Kindergarten test, which Minneapolis Public Schools administer to entering kindergartners.

[a] Fifteen of the 16 scholars in the Family Academy group also had been enrolled in high-quality early learning programs.

Opportunities and Resources

In addition to informing improvement of their own programs and strategies, evaluation presents Promise Neighborhoods with the opportunity to learn from, and add to, the body of evidence for place-based initiatives. To do this effectively, Promise Neighborhoods should be aware of the standards of evidence and models of evaluation established by other researchers. They should also consider basing their evaluations upon strong models to ensure their work both is informed by and, hopefully, can contribute to the current evidence base.

US Department of Education Evaluation Standards

The US Department of Education has set a hierarchy for describing how “evidence based” a given program is. This ranking system is defined in the Education Department General Administrative Regulations (or EDGAR).[12]

The EDGAR ranking allows practitioners to quickly identify which interventions have been shown to cause specific outcomes, based on rigorous summative evaluations. While recognizing that important insights are often available from other types of studies, the department’s rules place a premium on summative evaluations with positive findings. Even more specifically, the department’s highest standards for what is “evidence based” have two tiers, both of which require the application of summative evaluations methods.

- Tier 1, known as "strong evidence of effectiveness," is reserved for interventions that are rigorously evaluated by a high-quality randomized assignment study. This means that some students or families were randomly enrolled as program participants while others, also eligible for services, were excluded for the purposes of comparing outcomes. To qualify as a high-quality study, EDGAR standards specify that these groups must be drawn from a total of 350 or more students, and the programming must take place in at least two different sites.[13]
- Tier 2, or "moderate evidence of effectiveness," is assigned to interventions whose outcomes are supported by a high-quality study that involves a statistically controlled comparison group. In this tier, researchers can identify a comparison group with a wider range of methods known as quasi-experimental design. In addition to studies that involve a nonrandomized comparison group, this rating can be assigned to well-designed randomized studies that encountered significant attrition or other problems that affect their reliability. Although the criteria are more flexible, tier 2 methodologies have the same restrictions as tier 1 in requiring multiple sites and large study populations.

The next two EDGAR tiers represent lesser standards that are not considered to represent evidence-based outcomes.

- Tier 3, or "evidence of promise," is assigned to studies without statistical controls, such as correlation-based research.
- Tier 4, or "clear rationale," is assigned to programs that have articulated a plausible reasoning behind their intervention without necessarily having any evidence to confirm assumptions or results. Plausible reasoning should be expressed using a logic model articulating a theory of change.

Despite their designation in the EDGAR hierarchy, evidence tiers 3 and 4 play important roles for Promise Neighborhoods, as well as other programs doing place-based, community-level education work, particularly at the formative stages. Promise Neighborhoods must create new strategies to address long-enduring community disparities; a clear rationale and evidence of promise are often the right places to begin developing those solutions. Over time, more promising practices can be evaluated more rigorously.

What Works Collaborative

As part of evaluation planning, it is useful for practitioners to review the judgments of the What Works Clearinghouse (WWC), which is run by the Institute of Education Sciences. Particularly for summative evaluations, WWC has applied the Department of Education's evidence rating system to relevant studies and, in doing so, has made it relatively easy to see which evaluations have provided the most convincing evidence of program impact. While these standards are broadly aligned with the EDGAR framework, WWC emphasizes ranking summative evaluations that fit into evidence-based tiers 1 and 2.

WWC rankings in this process are particularly important for recent Promise Neighborhoods grantees. The 2017 regulations refer grantees to the WWC to help judge which studies are "high quality."[14] While criteria on sample size are articulated in the Promise Neighborhoods regulation, judgments relating to the balance of treated and untreated study participants, and which methods are appropriate for comparing across these groups, are more complex and nuanced. The WWC judges these study design choices in context, which is necessary for determining which evaluations are high quality.

Protection of Human Subjects

Researchers need to take appropriate steps to ensure that the rights and safety of individuals participating in evaluations or other studies are protected. Protection of human subjects in research involves such steps as obtaining informed consent before collecting data from or interviewing people, taking necessary precautions to safeguard personally identifiable or confidential information, and ensuring that any risks to participation are minimized. These protections ensure that research is conducted ethically.

Promise Neighborhood grantees also must comply with Department of Education regulations on human protections when conducting research, including evaluations.[15] These rules include having all covered research projects reviewed and approved by an Institutional Review Board (IRB) before the research takes place. IRBs are formal entities made up of a diverse group of people who can assess whether the proposed research will protect the rights and welfare of the people who participate. Organizations that do not have their own IRB can engage an independent one to perform this role.[16]

Conclusion

Evaluation is a vital component of a Promise Neighborhood's efforts to build an evidence base that can inform efforts to refine program implementation, improve results, and document the impact of individual programming and the overall cradle-to-career strategy. Though Promise Neighborhoods have distinct evaluation challenges, they can nonetheless learn from successful community-impact evaluations that have sought to answer similar questions.

As we reviewed other community impact models, one consistent theme we found was, wherever an evaluation effort culminated in a well-designed summative evaluation of a major program, it had started with formative evaluations. Based on these experiences, it seems that the challenges associated with community-based impact models often require iterative evaluations.

Further, answering the questions that Promise Neighborhoods need to address requires varied evaluation methods, including formative (planning and implementation studies) and summative (impact studies). Many Promise Neighborhoods may want to start with formative evaluations, with summative evaluations as something to aspire to. By adopting multiple approaches, Promise Neighborhoods can take full advantage of the range of evaluation methods and contribute to their own knowledge as well as to that of the larger field of place-based initiatives.

Notes

[1] See Emma Dorn, Bryan Hancock, Jimmy Sarakatsannis, and Ellen Viruleg, "COVID-19 and Student Learning in the United States: The Hurt Could Last a Lifetime," McKinsey & Company, June 1, 2020, https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-and-student-learning-in-the-united-states-the-hurt-could-last-a-lifetime; Sujata Gupta, "Creative School Plans Could Counter Inequities Exposed by COVID-19," *Science News*, September 8, 2020, https://www.sciencenews.org/article/coronavirus-covid-19-creative-school-plans-students-inequities; Garcia and Weiss (2020); Kuhfeld et al. (2020); and Vogels et al. (2020).

[2] To read profiles of all 28 Promise Neighborhoods grantees, go to https://promiseneighborhoods.ed.gov/data-and-results/infographics.

[3] The framework is available at https://promiseneighborhoods.ed.gov/data-and-results/results-framework.

[4] Application for New Awards; Promise Neighborhoods Program, 82 Fed. Reg. 33,881 (July 21, 2017).

[5] Urban Institute is a subcontractor to Mathematica for the national evaluation, providing guidance and advice on the evaluation approach.

[6] The feasibility study was included as an attachment to the national evaluation solicitation (form SF33, pages 69–177), which is available at https://beta.sam.gov/opp/a6beac5e677cac913c4959e7ef0e5456/view.

[7] Although they have some similarities, evaluation differs from performance management, which also uses data to assess program effectiveness. For a discussion of the differences, see Tatian (2016).

[8] "Process Evaluation vs. Outcome Evaluation," TSNE MissionWorks blog, June 14, 2018, https://www.tsne.org/blog/process-evaluation-vs-outcome-evaluation.

[9] Austin Nichols, "Evaluating Place-Based Programs," *Urban Wire* (blog), Urban Institute, July 3, 2013, https://www.urban.org/urban-wire/evaluating-place-based-programs.

[10] "Propensity Scores," The Methodology Center, Pennsylvania State University, College of Health and Human Development, accessed November 30, 2020, https://www.methodology.psu.edu/resources/propensity-scores/.

[11] Alberto Abadie, "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," article prepared for the *Journal of Economic Literature*, June 2020, available at https://economics.mit.edu/files/17847.

[12] Direct Grant Programs and Definitions That Apply to Department Regulations, 77 Fed. Reg. 42,928 (August 13, 2013) (to be codified at 34 C.F.R. pts. 75 and 77).

[13] If randomization is taking place at a larger group level, such as a classroom or school, there must be no fewer than 50 groups with at least 10 participants in each group.

[14] Application for New Awards; Promise Neighborhoods Program.

[15] See https://www2.ed.gov/policy/fund/guid/humansub/overview.html for an overview of the regulations surrounding protection of human subjects in research. Additional information is available at https://www2.ed.gov/rms-training/human-subjects-research/story.html.

[16] Grantees can search the Office for Human Research Protections database (http://ohrp.cit.nih.gov/search/irbsearch.aspx?styp=bsc) for registered IRBs.

References

Auspos, Patricia. 2012. *Developing and Using Data and Evidence to Improve Place-Based Work: Proceedings from a Meeting Convened by the Aspen Institute Roundtable on Community Change*. Washington, DC: The Aspen Institute

Bai, Rong, Cyleste C. Collins, David Crampton, Chun Lui, and Rob Fischer. 2017. "Cuyahoga Partnering for Family Success Process Evaluation 2017." Briefly Stated No. 17-01. Cleveland: Case Western Reserve University, Center on Urban Poverty and Community Development.

Black, Marsha M. 2004. *Measuring the Impact of HIPPY on Children: Pre-Post Testing.* Tampa: University of South Florida.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133 (3): 1,107–62.

Garcia, Emma, and Elaine Weiss. 2020. *COVID-19 and Student Performance, Equity, and U.S. Education Policy*. Washington, DC: Economic Policy Institute.

Giancola, Susan P. 2014. *Evaluation Matters: Getting the Information You Need from Your Evaluation*. Washington, DC: US Department of Education.

Johnston, William R., Celia J. Gomez, Lisa Sontag-Padilla, Lea Xenakis, and Brent Anderson. 2017. *Developing Community Schools at Scale: Implementation of the New York City Community Schools Initiative*. Santa Monica, CA: RAND Corporation.

Jutte, Douglas P., Kaja Z. LeWinn, Malo A. Hutson, Ramie Dare, and Janet Falk. 2011. "Bringing Researchers and Community Developers Together to Revitalize a Public Housing Project and Improve Health." *Health Affairs* 30 (11): 2,072–78. https://doi.org/10.1377/hlthaff.2011.07.01.

Kuhfeld, Megan, James Soland, Beth Tarasawa, Angela Johnson, Erik Ruzek, and Jing Liu. 2020. "Projecting the Potential Impacts of COVID-19 School Closures on Academic Achievement." Ed Working Paper 20-226. Providence, RI: Annenberg Institute for School Reform at Brown University. https://doi.org/10.26300/cdrv-yw05.

LFA Group. 2012. *HOPE SF: Baseline Evaluation Report*. San Francisco: Mayor's Office of Housing and Community Development.

Rockwell, S. Kay, and Harriet Kohn. 1989. "Post-Then-Pre Evaluation." *Journal of Extension* 27 (2).

Shelton, Ellen, Cael Warren, and Sarah Gehrig. 2017. *NAZ 2016 Annual Report: Including Accomplishments Over Five Years as a Promise Neighborhood*. Saint Paul, MN: Wilder Research.

Tatian, Peter A. 2016. *Performance Measurement to Evaluation*. Washington, DC: Urban Institute.

University of Wisconsin–Extension. 2005. "Quick Tips: Using the Retrospective Post-Then-Pre Design." Madison: University of Wisconsin–Extension.

Vogels, Emily A., Andrew Perrin, Lee Rainie, and Monica Anderson. 2020. "53% of Americans Say the Internet Has Been Essential during the COVID-19 Outbreak." Washington, DC: Pew Research Center.

About the Authors

Peter A. Tatian is a senior fellow at the Urban Institute and research director for Urban–Greater DC. He advises nonprofits on performance management and evaluation and co-leads Urban's work providing technical assistance to grantees of the US Department of Education's Promise Neighborhoods initiative. He directed the evaluation of the National Foreclosure Mitigation Counseling program, which provided counseling services to more than 1 million troubled homeowners during the foreclosure crisis, and has studied the impacts of public and supportive housing on neighborhoods.

Benny Docter is a research assistant in the Metropolitan Housing and Communities Policy Center. In both professional and academic contexts, Docter has built resident databases and used them to analyze long-term trends in tenancy. His research interests include affordable housing and neighborhood housing density. Before joining Urban, Docter worked in affordable housing development and community programming. He graduated with honors from Wesleyan University, majoring in economics and social studies.

Acknowledgments

This research was funded by the US Department of Education. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

The authors would like to thank Promise Heights Promise Neighborhood, Chula Vista Promise Neighborhood, and San Diego Promise Neighborhood for sharing their evaluation strategies and experiences. We also are grateful to Sara McTarnaghan and Elson Nash for reviewing and commenting on earlier versions of this brief.

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org