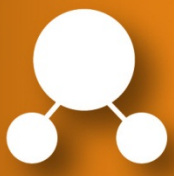
Meta-Analysis of Evaluations across the Social Innovation Fund Program: Final Report



ICF INTERNATIONAL

Corporation for NATIONAL & COMMUNITY SERVICE

A product of the National Assessment of the Social Innovation Fund (SIF)

SEPTEMBER 2016

Meta-Analysis of Evaluations across the Social Innovation Fund Program: Final Report

September 2016

Authors

Xiaodong Zhang, Ph.D.
Jing Sun, M.A.

Submitted To

Lily Zandniapour, Ph.D.
Office of Research and Evaluation
Corporation for National and Community Service
For inquiries, please contact evaluation@cns.gov

Acknowledgements

Janet Griffith, Ph.D and Janet Pershing, Ph.D., of ICF International, provided review and insights that informed the development of this product.

About This Document

This document was produced pursuant to a contract between the Corporation for National and Community Service and ICF International (Contract Number: GS-23F-8182H; CNSHQ13F0091).

ICF International | 9300 Lee Highway | Fairfax, VA 22031-1207 | www.icfi.com

Recommended Citation

Zhang, X. & Sun, J. (2016). *Meta-Analysis of Evaluations across the Social Innovation Fund Program: Final Report*. (Prepared for the Corporation for National and Community Service, Office of Research and Evaluation). Fairfax, VA: ICF International.

About CNCS

The Corporation for National and Community Service (CNCS) is a federal agency that improves lives, strengthens communities, and fosters civic engagement through service and volunteering. Each year, CNCS engages five million Americans of all ages and backgrounds through AmeriCorps, Senior Corps, the Social Innovation Fund, the Volunteer Generation Fund, and other programs, and leads the President's national call to service initiative, United We Serve. For more information, visit NationalService.gov.

Table of Contents

Executive Summary

The Corporation for National and Community Service (CNCS) implements the Social Innovation Fund (SIF), an initiative that combines public and private resources to grow the impact of innovative, evidence-based solutions to improve the lives of people in low-income communities throughout the United States. As part of the national assessment of SIF,[1] ICF International conducted a meta-analysis of 38 final evaluation reports completed by evaluators for the SIF grantees and subgrantees and submitted to CNCS in order to assess the state of SIF evaluation findings as of May 2016. Meta-analysis is a rigorous and cost-effective approach to synthesizing a large body of research and evaluation findings. The findings from this report are mostly from the 2010 cohort, as these groups of grantees or subgrantees had completed their evaluations. Targeting and achieving moderate or strong evidence was not required of grantees in the 2010 cohort. For later cohorts, SIF's oversight and support of grantee and subgrantee evaluations evolved significantly. It is anticipated that a greater proportion of the evaluations in later cohorts will meet the criteria for strong or moderate levels of evidence. Below is a summary of main findings and conclusions.

The SIF catalyzes innovation in targeted sectors. Thirty-two percent of the reports were in the healthy futures issue area, 45% in youth development (including early childhood), and 24% in economic opportunity. Funded programs employed a variety of programmatic strategies for the target populations they served. Examples include summer programs, one-on-one tutoring, social enterprises, workforce partnerships, community-based health care, and telemedicine. All the programs had multiple key components, consistent with the expectation that addressing social issues requires complex and multi-layered solutions.

SIF evaluations employed rigorous methods to assess program implementation and impact. Building a sound evidence base supported by rigorous evaluations is critical to the SIF. The evaluations reviewed included both implementation evaluations and impact/outcome evaluations. Overall, 32 of the 38 reports evaluated implementation, and 32 evaluated outcomes or impacts (26 reported on both). More than half (56 %) of the 32 impact/outcome evaluations employed such rigorous designs as randomized controlled trial (RCT) or quasi-experimental (QED) with matching, which allows causal inferences. Half of the impact/outcome evaluations appeared to be adequately powered in all or at least one outcome. Most of these evaluations used valid and reliable outcome measures to capture the expected outcomes of interventions, and employed inferential statistics to support the findings.

Overall, of the 38 final evaluation reports, 5 were categorized as having produced a strong level of evidence, 8 as providing a moderate level, and 25 (including 6 with implementation only results) as providing a preliminary level. The level of evidence refers to the rigor of the evaluation rather than whether or not the evaluation generates positive findings. The levels are based on how well a particular evaluation addresses concerns about internal and external validity, with more effective evaluations categorized as attaining strong or moderate levels, and less effective evaluations as attaining preliminary levels of evidence.

Early findings show that SIF interventions generated meaningful and statistically significant impact overall and across sectors. The evaluation findings show that most of the SIF-funded interventions (87%) for which evaluation reports are available were implemented with fidelity. The same percentage of the programs were found to have had a positive impact on all or some target outcomes.

Meta-analysis of impact findings from 28 impact/outcome reports that provide sufficient statistics show statistically and practically significant effects. The average effect size was 0.28 and statistically

[1] The national assessment covers SIF Classic grants; it does not cover SIF Pay for Success grants.

significant. This can be interpreted as an 11 percentile gain by the treatment over the comparison group. The average effect is the largest in economic opportunity intervention (0.40), followed by healthy futures (0.32) and youth development (0.08), all of which were statistically significant. We characterized outcomes evaluated by the evaluation reports into 13 broad outcome domains so that we could group similar outcome metrics under the same domains (e.g., smoking, physical activity, alcohol use under "health behavior"). Results from 10 out of 13 outcome domains were statistically significant at the 0.10 level. The non-significant findings were partly due to small sample sizes. Finally, effect sizes from studies with preliminary evidence is 0.26, moderate evidence is 0.16, and strong evidence is 0.48, all of which were statistically significant.

SIF evaluations advanced knowledge about program implementation and evaluation in substantive ways. Close to half the reports reflected on challenges, limitations to the findings, and lessons learned. For example, program-related lessons included suggested improvements to program elements and to working relationships among stakeholders, and other practices related to participant recruitment and retention, length of intervention, modest start-up, cost implications, infrastructure, and expansion to different populations. Evaluation-related lessons concerned adequate funding for data collection, the value of sizeable administrative datasets, tracking longitudinal outcomes, and others, such as statistical power, detail about the control condition, and problems posed by selection of inexperienced sites.

Meta-analysis provides a rigorous and efficient tool to synthesize findings across evaluations. Evidence generated by SIF-funded interventions is likely to become one of the SIF's enduring legacies and will inform future innovation and scale-up by other nonprofits across the nation. These evaluations contribute to the social innovation knowledge base in the fields of child/youth development, economic opportunity, and community health, thus complementing the evaluation efforts of other federal agencies such as the U.S. Departments of Education, Labor, and Health and Human Services, as well as nonprofit and philanthropic organizations. The SIF has a fully operational pipeline of on-going evaluations that will generate more results in the coming years. It is reasonable to expect that as results from more SIF-funded evaluations become available, they can be summarized using meta-analysis to further advance the knowledge base. The evidence generated from the SIF will contribute significantly to the evidence base from other funders working in the same areas.

Findings from this report have important implications for the stakeholders.

- **For policymakers.** Statistically and practically significant results across evaluations validate the significant CNCS investment in evaluating different program models designed to improve a wide range of outcomes and increase their impact. As more evaluations are completed, policymakers should continue to ensure that the findings are synthesized and disseminated in order to develop the knowledge base and distill these lessons into accessible guidance for nonprofits seeking to implement evidence-based interventions.
- **For philanthropy and nonprofits.** The findings demonstrate the power of rigorous evaluation. Practitioners should recognize the importance not only of providing services to populations in need, but also of building evidence to inform replication and scaling up for wider implementation. These evaluations provide valuable lessons for improving program implementation, including cultivating relationships among stakeholders, articulating programmatic strategies and theories of change, considering cost implications when making program decisions, and recognizing the importance of adaptation, flexibility, and individualization in scaling effort. They also highlight the importance of developing the practices of nonprofit organizations to promote evidence-based programs and evaluations.
- **For evaluators.** SIF provides an opportunity and funding to test the efficacy of innovative programs that have at least preliminary evidence of effectiveness. The meta-analysis approach illustrates the contribution these individual evaluations can make to growing the evidence base. These evaluations

also provide crucial lessons for strengthening evaluation, such as exploring the full range of design options in light of questions of interest, rigor, and feasibility; recognizing the importance of conducting an implementation study as well as an impact study, and thinking through all aspects of data collection—including data access, sampling, recruitment, and retention—to anticipate potential challenges and develop solutions.

1. Introduction

The Corporation for National and Community Service (CNCS) implements the Social Innovation Fund (SIF), a program that combines public and private resources to grow the impact of innovative, evidence-based solutions to improve the lives of people in low-income communities throughout the United States. The program was launched in 2010 following the implementation of the Serve America Act, and is one of six Obama Administration "tiered-evidence initiatives" embodying the principles of social innovation. To date, the SIF has received an annual Congressional appropriation of approximately $50-70 million. The program leverages federal funds through public-private collaborations by granting money to intermediary grantmakers who find, fund, improve, and grow promising community-based solutions with evidence of successful outcomes in three core areas: healthy futures, youth development, and economic opportunity.

- Youth development: Preparing children and youth for success in school, active citizenship, productive work, and healthy and safe lives, including crime reduction initiatives focused on juvenile delinquency and victimization prevention and response.
- Economic opportunity: Increasing economic opportunities and financial stability for economically disadvantaged individuals and families.
- Healthy futures: Improving health outcomes, promoting healthy lifestyles, and decreasing health disparities that disproportionately affect low-income communities.

As described by CNCS, the SIF is characterized by the unique interplay of six key elements. The program:

1) relies on intermediary grantmaking institutions to implement the program–these organizations take on the role of finding, selecting, monitoring, supporting, evaluating, and reporting on the nonprofit organizations they fund to implement community-based interventions.
2) is a tiered-evidence initiative that requires all funded programs/interventions to demonstrate at least preliminary evidence of effectiveness, or funding "what works."
3) requires that all programs or interventions implement a rigorous independent evaluation that will build on their initial level of evidence.
4) charges intermediaries with scaling evidence-based programs—increasing impact within their community or to communities across the country—and grapples with the challenge of how to do so successfully and efficiently.
5) leverages public-private partnerships to effect large-scale community impact in ways that neither a traditional federal grant nor a philanthropic investment could achieve on its own. The program requires a dollar for dollar match at the grantee level, and then dollar for dollar match at the subgrantee level.
6) is committed to improving the effectiveness of nonprofits, funders, and other federal agencies by capturing and learning about best practices and promoting approaches that will generate the greatest impact for individuals and communities.

To date, the SIF program has implemented six rounds of grantmaking through its "Classic" grant program (in 2010, 2011, 2012, 2014, 2015, and 2016) and selected 33 organizations to receive 39 grants. (Several organizations received multiple awards to implement different initiatives or expand the same initiative.). These intermediaries, herein referred to as SIF grantees, have in turn selected about 300 nonprofit organizations through open and competitive processes to implement promising interventions. (The 2015 and 2016 SIF grantees are still in the process of selecting subgrantees.) SIF grants are for $1 million to $10 million per year, for five years. Subgrants are $100,000 or higher for a period of 3-5 years.

It appears that the 2010 cohort grantees have funded more subgrants than did later cohorts. Exhibit 1 summarizes the number of awards by year.

Exhibit 1. *SIF Grantees and Subgrantees by Grantee Award Year*

Award year	Number of Grants to SIF grantees	Number of SIF subgrantees funded
2010	11	149
2011	5	48
2012	4	15
2014	7	40
2015	8	TBD
2016	4	TBD
Total	35	~300*

Extracted from CNCS website.
*This is an approximate total as 2015 grantees are still in the process of selecting subgrantees.

As a federal tiered-evidence initiative, the SIF is committed to assessing the rigor level of evidence, both to select grantees for funding and to evaluate the outcomes/impacts of the program models. To achieve this goal, CNCS expects SIF grantees and subgrantees to enter the program with interventions that have at least preliminary evidence of effectiveness *and* to commit significant time and resources to conducting formal evaluations of program models that receive SIF funding. The SIF program office, together with the staff of the CNCS Office of Research and Evaluation (ORE), as well as the evaluation technical assistance contractor and consultants, work closely with grantees and subgrantees by providing them with technical assistance in designing, implementing, and monitoring the SIF Evaluation Plan (SEP),[2] collecting best practices to share with the broader social innovation sector, and disseminating evidence of effectiveness for each program model within the SIF. The tiered-evidence framework is presented in Exhibit 2 (CNCS, n.d.).

Exhibit 2. *SIF Tiered-Evidence Framework*

- Preliminary evidence means that the model has evidence, based on a reasonable hypothesis and supported by credible research findings. Examples of research that meet this standard include: 1) outcome studies that track participants through a program and measure their responses at the end of the program; and 2) third-party pre- and post-test research that determines whether participants have improved on an intended outcome.
- Moderate evidence means evidence from previous studies with designs that support causal conclusions (i.e., studies with high internal validity) but limited generalizability (i.e., moderate external validity) or vice versa—studies that only support moderate causal conclusions but have broad general applicability. Examples of studies that would constitute moderate evidence include: 1) at least one well-designed and well-implemented experimental or quasi-experimental study supporting the effectiveness of the practice strategy or program, with small sample sizes or other conditions of implementation or analysis that limit generalizability; or 2) correlational research with strong statistical controls for selection bias and for discerning the influence of internal factors. Moderate evidence requires third-party or external and impartial evaluators.
- Strong evidence means evidence from previous studies with designs that support causal conclusions (i.e., studies with high internal validity), which, taken together, include enough of the range of participants and settings to support scaling up to the state, regional, or national level (i.e., studies with high external validity). Examples of studies that would qualify as strong evidence include: 1) more than one well-designed and well-implemented experimental study or well-designed and well-implemented quasi-experimental study that supports the effectiveness of the practice, strategy, or program; or 2) one large, well-designed and well-implemented randomized controlled, multisite trial that supports the effectiveness of the practice, strategy, or program. Strong evidence requires third-party or external and impartial evaluators.

As part of the national assessment of the SIF, ICF International has conducted this meta-analysis of evaluations prepared by SIF grantees and subgrantees to capture the state of evaluation findings from

[2] Each SIF grantee/subgrantee evaluator is required to prepare a SEP that details the program model to be evaluated and justifies the evaluation approach selected.

the SIF. Meta-analysis is a rigorous and cost-effective approach to synthesizing a large body of research and evaluation findings. The cross-evaluation analysis focuses on the following research questions:

- What are the key programmatic features of the SIF-funded programs?
- What are the methods used to evaluate the SIF-funded programs?
- What are the evaluation findings from the SIF-funded programs?

This report presents findings from 38 final evaluation reports completed by grantees or subgrantees, submitted to and reviewed by CNCS as of May 1, 2016.

2. Methodological Approach

This chapter describes the methodological approach used in this report, including strategies for identifying evaluation reports and synthesizing findings across studies. More detail on the methodology is presented in Appendix A.

2.1 Final Evaluation Reports Reviewed

A total of 38 final evaluation reports were included in this meta-analysis, after their review by CNCS and its technical assistance contractor or consultants (Exhibit 3). Reports completed by evaluators of grantees/subgrantees but still under review were not included in this analysis. According to a recent CNCS report (CNCS, 2015), the SIF has received a total of 108 SEPs covering 87 programs, with 77 plans approved. The approved SIF Evaluation Plans include both randomized controlled trials (experimental studies) and quasi-experimental studies.

The 38 reports include 26 synthesized in a report titled: *Meta-synthesis of Evaluations Across Social Innovation Fund (SIF): Interim Report* (Zhang, 2015) and 12 additional reports finalized between June 1, 2015 and May 1, 2016. Each report describes findings from a unique intervention. A SIF grantee may implement one unique intervention across multiple subgrants and by extension, use one evaluation plan for the program (referred to as a UniSEP), or multiple interventions across the subgrants within its portfolio, which would be evaluated using multiple evaluations (referred to as a MultiSEP). Four out of the 38 are UniSEP, whereas 34 are MultiSEP. The 38 reports come from programs funded through the first two SIF cohorts: 9 grantees with 28 subgrantees from the 2010 cohort and 3 grantees with 6 subgrantees from the 2011 cohort. Four report findings at the grantee-level, and the rest report findings at the subgrantee level.

It is important to note that, until the 2011 cohort, the SIF had not clearly articulated the requirement that grantee evaluations be designed to target and hopefully achieve a moderate or strong level of evidence. As such, the SIF permitted the 2010 grantees and subgrantees to conduct evaluations with a range of targeted evidence levels, including preliminary level of evidence, so long as they were building on the existing base of evidence for the intervention and advancing it in a meaningful way. As a result, although all the 2010 cohort grantees and subgrantees conducted evaluations, not all conducted impact evaluations and, among those that did, some did not design evaluations to achieve a moderate or strong level of evidence, due to funding limitations, data restrictions, or other factors. Additionally, the experience of the initial 2010 SIF cohort appears, based on anecdotal evidence, to differ in other important ways from that of later cohorts. For example, many SIF grantees and subgrantees in the first cohort entered the program with little understanding of the SIF's evidence expectations or of related requirements such as Institutional Review Board (IRB) reviews (Lester, 2015).

To the extent that the mix of organizations or the experience of later cohorts reflects the effects of more rigorous evaluations, the ability to generalize findings from this report to the larger population of later cohorts and future SIF grantees is limited. For later cohorts, SIF's oversight of and support to grantee and subgrantee evaluations evolved significantly. It is anticipated that a greater proportion of the evaluations in later cohorts will meet the criteria for strong or moderate levels of evidence.

Exhibit 3. *SIF Final Evaluation Reports Completed and Reviewed by May 1, 2016*

Cohort	Grantee	Subgrantee	Evaluation organization	Program name	Priority area	Level of evidence
2010	Edna McConnell Clark Foundation	Building Educated Leaders for Life (BELL)	MDRC	BELL Middle School Model	Youth development	Moderate
2010	Edna McConnell Clark Foundation	Center for Employment Opportunities	MDRC	Center for Employment Opportunities (CEO)	Economic opportunity	Preliminary
2010	Edna McConnell Clark Foundation	Gateway to College National Network	MDRC	Gateway to College	Youth development	Preliminary
2010	Edna McConnell Clark Foundation	Reading Partners	MDRC	Reading Partners	Youth development	Strong
2010	Edna McConnell Clark Foundation	SEED Foundation	MDRC	SEED School	Youth development	Moderate
2010	Edna McConnell Clark Foundation	Communities in Schools (CIS)	MDRC	CIS	Youth development	Strong
2010	Edna McConnell Clark Foundation	Children's Aid Society	Philliber Research and Evaluation	Carrera Adolescent Pregnancy Prevention Program (CAS-Carrera)	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	Cumberland Family Medical Center, Inc.	University of Kentucky	Cumberland Family Medical Center	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	Home of the Innocents	REACH Evaluation	Open Arms Children's Health program	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	King's Daughter Medical Center	University of Kentucky, Prevention Research Center	Mobile Health Services for Rural Kentucky (MHSRK) program	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	Meade Activity Center	University of Louisville	Meade Activity Center	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	Montgomery County Health Department	CHES Solutions Group	Community Health Worker (CWH) program	Healthy futures	Preliminary

Cohort	Grantee	Subgrantee	Evaluation organization	Program name	Priority area	Level of evidence
2010	Foundation for a Healthy Kentucky	Norton Health System GHAP	University of Kentucky, Prevention Research Center	Get Healthy Access Program (GHAP)	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	Oldham County Health Dept./Hope Health Clinic	University of Louisville	Hope Health Clinic	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	St. Elizabeth	Innovations	St. Elizabeth Telepsychiatry program	Healthy futures	Preliminary
2010	Foundation for a Healthy Kentucky	St. Joseph Health System	St. Joseph Health System	Community Based Delivery Model: Virtual Care	Healthy futures	Preliminary
2010	Jobs for the Future	National Fund for Workforce Solutions UniSEP	IMPAQ International	Workforce Partnership Programs	Economic opportunity	Moderate
2010	Local Initiatives Support Corporation (LISC)	LISC	Economic Mobility Corporation	Financial Opportunity Centers (FOC)	Economic opportunity	Moderate
2010	Mayor's Fund	Bronx Works	MDRC	Jobs-Plus	Economic opportunity	Preliminary
2010	Mayor's Fund	Project Rise	MDRC	Project Rise	Economic opportunity	Preliminary
2010	Mayor's Fund	SaveUSA	MDRC	SaveUSA	Economic opportunity	Strong
2010	Mayor's Fund	WorkAdvance	MDRC	WorkAdvance	Economic opportunity	Strong
2011	Mile High United Way	Colorado Parent and Child Foundation	University of Denver	Parents as Teachers (PAT), and Home Instruction for Parents of Preschool Youngsters (HIPPY)	Youth development	Preliminary
2011	Mile High United Way	Jefferson Foundation	APA Consulting	Jeffco Summer of Early Literacy (JSEL)	Youth development	Preliminary
2010	Missouri Foundation for Health	Social Innovation for MO UniSEP	Washington University of St. Louis, Center for Tobacco Policy Research	Social Innovation for Missouri	Healthy futures	Preliminary
2010	REDF	REDF UniSEP	Mathematica	Social Enterprises	Economic opportunity	Moderate
2010	United Way of Greater Cincinnati	Cincinnati Arts and Technology Center	Innovations	Bridging the Gap	Youth development	Preliminary
2010	United Way of Greater Cincinnati	Cincinnati Museum Center	Innovations	Early Childhood Science Inquiry Training for Educators (ECSITE)	Youth development	Preliminary

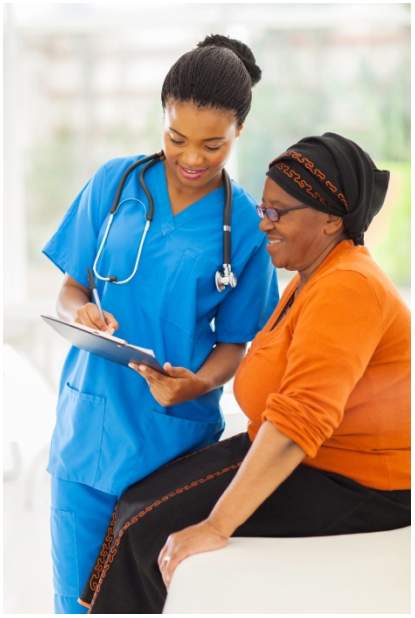
Cohort	Grantee	Subgrantee	Evaluation organization	Program name	Priority area	Level of evidence
2010	United Way of Greater Cincinnati	Covington Public Schools	University of Cincinnati, Evaluation Services Center	Holmes 180	Youth development	Preliminary
2010	United Way of Greater Cincinnati	Easter Seals	University of Cincinnati, Evaluation Services Center	Transitional Employment Services Model	Economic opportunity	Preliminary
2010	United Way of Greater Cincinnati	Resilient Children Project	University of Cincinnati, Evaluation Services Center	Resilient Children Project	Youth development	Preliminary
2010	United Way of Greater Cincinnati	University of Cincinnati	University of Cincinnati, Evaluation Services Center	UC Degrees Gen-1 Project and Higher Education Mentoring Initiative	Youth development	Preliminary
2011	U.S. Soccer Foundation	USSF UniSEP	U.S. Soccer Foundation, Healthy Networks Design and Research, Child Trends	Soccer for Success	Healthy futures	Strong
2011	United Way for Southeast Michigan	Detroit Public Television	Michigan Public Health Institute	Pre-School-U	Youth development	Preliminary
2011	United Way for Southeast Michigan	Detroit Parent Network	Evaluation Strategies	Pathway to Literacy (PTL)	Youth development	Moderate
2011	United Way for Southeast Michigan	Starfish Family Services	Public Sector Consultants	Inkster Family Literacy Movement	Youth development	Preliminary
2011	United Way for Southeast Michigan	Guidance Center	Innovatus Consulting	Families and Schools Together (FAST); Community Organizing Around Family Issues (COFI)	Youth development	Preliminary
2010	Venture Philanthropy Partners	Latin American Youth Center	Urban Institute	Promotor Pathway Program	Youth development	Moderate

*SIF program has a 3-1 matching requirement for federal dollars.
Source. CNCS website.

2.2 Review, Synthesis, and Analysis Methods

Meta-analysis is a rigorous and cost-effective approach to synthesizing a large body of research and evaluation findings. The method has taken on increasing policy importance. Compared to evaluations that use a singular design and rely on primary data collection, meta-analysis provides a more flexible, cost-effective, and less burdensome approach. This is especially true for a complex program such as SIF, where the heterogeneity of its projects makes it impossible and inappropriate to use a singular research

design such as RCT. The meta-analysis has two components: 1) thematic synthesis of the interventions and associated evaluations, and 2) estimation of the magnitude of program impact using effect sizes. Specifically, we reviewed and synthesized information from the final evaluation reports (listed in Exhibit 3) and incorporated findings from a systematic assessment of their quality by the evaluation TA contractor and consultants. We also reviewed other CNCS reports and external sources to contextualize the findings.



Each report was coded according to the three research questions (programmatic features, evaluation methods, and evaluation findings), sorting the information further into sub-topics. Next, emerging themes were summarized across reports and each illustrated with examples. When possible, these examples were drawn across the SIF's three priority areas. Finally, effect sizes were calculated based on statistics in the reports. In order to provide more parsimonious results, outcomes were grouped under 13 broad outcome domains. A total of 72 effect size estimates in 13 outcome domains were extracted from 275 effect sizes in the evaluation reports (See section 5.3 for details).

For the synthesis of themes, the unit of analysis is the final evaluation report, regardless of whether it concerns a grantee or subgrantee, because each report targets a unique program model or intervention. The sample size is 38 reports. In discussing the substantive findings, the reports are identified by these program names rather than as grantee or subgrantee reports. However, for effect size transformation, the unit of analysis is the effect size of an outcome domain from an independent sample in the evaluation reports. The sample size is 72 effect size estimates from the 13 outcome domains.

The next three chapters summarize and analyze findings across the reports according to the three research questions or categories (programmatic features, evaluation methods, and findings), and then sort each category further into sub-topics in each chapter.

3. Programs Features

Information about the characteristics of the funded programs and reports helps describe the nature of the interventions, providing the context needed to understand the evaluation methods and interpret the findings. In this section, we present data about the grantees, subgrantees, and evaluation organizations; outcome areas and target populations; programmatic strategies; key program components; and numbers of individuals served.

3.1 Grantees, Subgrantees, and Evaluation Organizations

SIF grantees employ two types of sub-award funding strategies. Some identify an evidence-based intervention upfront, and these grantees have typically used a single evaluation plan (UniSEP). The second group of SIF grantees invest their sub-award funds in organizations that have developed or implemented a variety of interventions to address one or more specific problems, and these grantees typically used a number of evaluation plans (MultiSEP) to test the supported interventions. As illustrated in Exhibit 3, independent external evaluators prepared the 38 evaluation reports on behalf of various configurations of grantees and subgrantees.

Exhibit 3 also shows that the evaluators for these reports came from a wide range of backgrounds. Twenty-five (66%) reports were authored by external research and evaluation organizations; 11 (29%) by universities or affiliated units; and 2 (5%) by the subgrantee itself. Some evaluators were responsible for more than one evaluation report.

3.2 Grantee Resources

A major factor influencing the success of these evaluations is the availability of resources sufficient to meet SIF's challenging evaluation demands. These resources include the pre-existing evaluation capacity of grantees and subgrantees, and the amount of evaluation funding relative to the evaluation scope. As we interpret the findings, it is important to keep in mind the varying resource levels associated with the grants, subgrants, and evaluations. A recent report (Lester, 2015) categorized SIF grantees and subgrantees into three broad categories:

- **Intermediaries with large grants.** A few organizations—such as the Edna McConnell Clark Foundation, the Mayor's Fund to Advance New York City, Financial Opportunity Centers, and Jobs for the Future—received the largest grants, with total federal funding of $12-30 million and substantial evaluation budgets;
- **Grantees with unified portfolios.** Although nonprofits such as REDF, Local Initiatives Support Corporation, Venture Philanthropy Partners, and the U.S. Soccer Foundation received smaller grants, they achieved efficiencies due to the fact that they were replicating one model in various locations rather than funding multiple interventions;
- **Small, regional intermediaries supporting multiple interventions.** Grantees such as Foundation for a Healthy Kentucky, United Way of Greater Cincinnati, and United Way for Southeast Michigan, in early cohorts typically received SIF grants at or near $1 million per year and provided subgrants at or near the $100,000 annual minimum. This funding structure produced a relatively large number of subgrantees with relatively little funding and therefore smaller evaluation budgets.

Of the reports evaluated for this synthesis, 14 (37%) were submitted by intermediaries with large grants; 3 (8%) by grantees with unified portfolios; and 21 (55%) by grantees that funded numerous, smaller, regional subgrants supporting different interventions. We observed correlations between grantee resources and level of evidence generated. According to Exhibit 3, 8 of the 13 reports completed by intermediaries with larger grants were rated as contributing to strong or moderate evidence; all 4 reports by grantees with unified portfolios were rated as strong or moderate; and only 1 of the 21 reports by grantees that funded smaller subgrants supporting multiple interventions were rated as strong or moderate.

A CNCS report (Zandniapour and Vicinanza, 2013) pointed out that many SIF grantees and subgrantees faced the challenge of budgeting adequately for evaluation. Exhibit 4 presents information about program and evaluation budgets for 70 interventions supported by the SIF. The same report showed that stronger levels of evidence come at a cost. Among the SIF evaluations, the average cost of randomized controlled trials (RCTs) per year is almost four times that of quasi-experimental (QEDs) or non-experimental designs. RCTs are not necessarily more costly, however, as it can be more difficult and more expensive to execute a well-designed non-experimental study, due to the challenge of establishing a strong, credible counterfactual. In addition, there are sound strategies for lowering the costs of RCTs.

Exhibit 4. Annual Budget for the SIF Grantee/Subgrantee Program and Evaluation

	Minimum	Maximum	Mean	Median
Program Budget	$100,000	$5,460,618	$1,104,649	$593,309
Evaluation Budget	$12,000	$1,346,342	$216,838	$81,471
Evaluation-to-Program Budget Ratio	12%	25%	20%	14%

Source: Zandniapour and Vicinanza (2013)

3.3 Target Populations and Outcome Areas

Of the 38 programs for which evaluation reports were available, 17 (45%) were in youth development, 9 (24%) in economic opportunity, and 12 (32%) in healthy futures. The target populations served and the outcomes the programs sought to achieve included the following:

- **Youth development.** All 17 programs served needy and at-risk populations, defined by low income or low achievement. Seven of the programs were in early childhood education, 8 in K-12 education, and 2 in postsecondary education. These interventions were designed to improve outcomes that included academic achievement, engagement, school readiness, school attainment (e.g., course enrollment, graduation, and retention), job readiness, life skills, and employment.
- **Economic opportunity**. All 9 programs were designed to serve economically disadvantaged populations, including individuals or families that were high school drop-outs, unemployed, hard-to-employ, low-income, on parole, homeless or on housing assistance, or contending with health issues. These programs targeted multiple outcomes, including job readiness, job skills, certifications, employment, job retention, earnings, postsecondary enrollment, financial well-being, and life stability outcomes such as stable housing, reduced criminal recidivism, and improved health.
- **Healthy futures**. All 12 programs served low-income and uninsured populations with 8 targeting rural areas, 2 targeting urban areas, and 2 targeting generally underserved communities. These programs aimed to improve various outcomes such as health care utilization, health behavior, satisfaction with health care, and health outcomes.

3.4 Programmatic Strategies

The evaluation reports describe the programmatic strategies employed by grantees and subgrantees to tackle the social problems and reach the target outcomes. These strategies, summarized below, varied by SIF priority area. Appendix B provides additional details about programmatic strategies.

- **Youth development**. Programs employed a variety of models, including summer programs, dual-enrollment (i.e., high school students taking college courses), one-on-one tutoring, mentorship, inquiry-based curriculum, professional development, comprehensive school reform, residential programs, training for parents, charter schools, caregiving on site or through home visits, and community partnership.
- **Economic opportunity**. Programs involved transitional jobs programs, workforce partnership, training/financial incentives/community support, social enterprises, work experience and skill training, financial counselling, and savings programs.
- **Healthy futures**. Programs used various approaches to increase access to and quality of health care, including patient navigation systems, mobile or telemedicine, community-based care, school-based prevention programs, after-school programs, and mobilizing support from a variety of stakeholders (e.g., volunteer doctors, faculty members, or community care workers).

3.5 Components of Interventions

Programmatic strategies are put into practice through concrete interventions, and those interventions usually consist of several different components. The implementation of multiple components is consistent with the recognition that the complex and layered problems tackled by SIF grantees require complex and multi-layered solutions.

One of the challenges evaluators face is the question of how to account for participants who do not receive all of the key components of a planned intervention. For example, some patients in a SIF-sponsored health initiative may be exposed to only one service, or may leave the program before services are completed. This not only reduces the intervention's effectiveness for that individual, but also complicates efforts to evaluate the effectiveness of the entire program.

Below are examples that illustrate how grantees combine multi-layered solutions within a priority area.

- **Youth development**. University of Connecticut (UC) Degrees Gen-1 Project and Higher Education Mentoring Initiative provides a variety of services, including social-emotional support (highly structured/supportive housing environment with curfews and mandatory house meetings), student academic support (house classes, academic conferences, studying/tutoring, time management guidance, and "intrusive" advising), community building (ceremonies, awards, events, celebrations, and retreats), and career development.
- **Economic opportunity**. Jobs Plus is an employment program designed to improve the economic well-being of public housing beneficiaries through employment and training services (job placements, job search training, and GED/ESL programs) as well as financial incentives (rent-based incentives like the Earned Income Disallowance program), and community support (for example, employing a subset of residents as community coaches to raise program awareness and buy-in).
- **Healthy futures**. Mobile Health Services for Rural Kentucky has three components: 1) implementing a mobile service to provide free health screenings and education to the target population; 2) providing fee-based cardiac testing for individuals found to be at-risk during screenings; 3) referring individuals screened and tested who need additional care to physicians for follow-up care.

3.6 Number of Individuals Served

The number of individuals served by the SIF projects ranged widely, from 76 to 7,500 with a mean of 1,480.[3] Several caveats are worth noting in interpreting these numbers:

- The scope of funding among grants and subgrants varied considerably. The annual SIF subgrantee awards ranged from $100,000 to $2 million;
- The nature and intensity of services differed across projects. For example, a medical program may serve a large number of patients only once, while an intensive training program for caregivers may work in-depth with a small number of participants;
- Although most programs targeted a single population, several worked with multiple populations. For example, the Resilient Children Project combined work with both teachers (280) and children (2,000) to improve school readiness by integrating education and health services; and

[3] This computation excludes outliers from Social Innovation for Missouri, which was designed to effect policy changes in seven counties, potentially reaching a population of 536,737, and Communities in Schools, which serves 53 schools.

- Because the SIF is designed to build an evidence base and to support present and potentially future scale-up, serving many individuals may be less important in the short term than gathering solid evidence about the program's implementation, outcomes, and impact.

4. Methods to Evaluate the Programs

The SIF provides resources and support for developing and implementing rigorous evaluations with the goal of building the evidence base of funded interventions as they expand their impact. Although the SIF has multiple goals, including innovation, evidence, and scale, Lester (2015, p. 9-10) argues that "SIF's primary impact comes not from its ability to scale effective programs within the confines of a very small federal program housed in CNCS, but its ability to build the evidence base and to show how to do so effectively, using very specific strategies rooted in venture philanthropy. Evidence trumps scale."

SIF grantees conducted a combination of implementation, impact, and outcome evaluations. Impact evaluations tend to use more rigorous designs, such as RCTs or QEDs, which involve counterfactuals that allow for attribution. In contrast, outcome evaluations tend to use cohort designs, pre/posttest designs or posttest only designs, which cannot demonstrate a causal relationship between program participation and the observed changes in outcomes. This report groups impact and outcome evaluations together. Although grantees and subgrantees in later SIF cohorts were required to conduct impact evaluations, those in the 2010 cohort, which represent 84% of the evaluation reports synthesized for this report, were not.

Exhibit 5 shows the implementation and impact/outcome focus of the reports examined in this synthesis. Note the aggregate numbers of reports that addressed each type of evaluation: 32 of the 38 reports (84%) provided evidence about program implementation, while 32 (84%) provided evidence about program impact or outcomes, and 26 (68%) addressed both implementation and impact.

Exhibit 5. *Focus of the Evaluation Reports*

	Total (N=38)		Aggregate Implementation/ Outcome Reports Combined with Single Focus Reports	
Report focus	n	%	n	%
Joint Focus: Implementation and impact/outcome	26	68%		
Single Focus: Implementation	6	16%	32	84%
Single Focus: Impact/outcome	6	16%	32	84%

In this chapter, we discuss methods that grantees and subgrantees used to evaluate the implementation and impact/outcome of the funded programs, examining a number of important (but by no means exhaustive) topics.

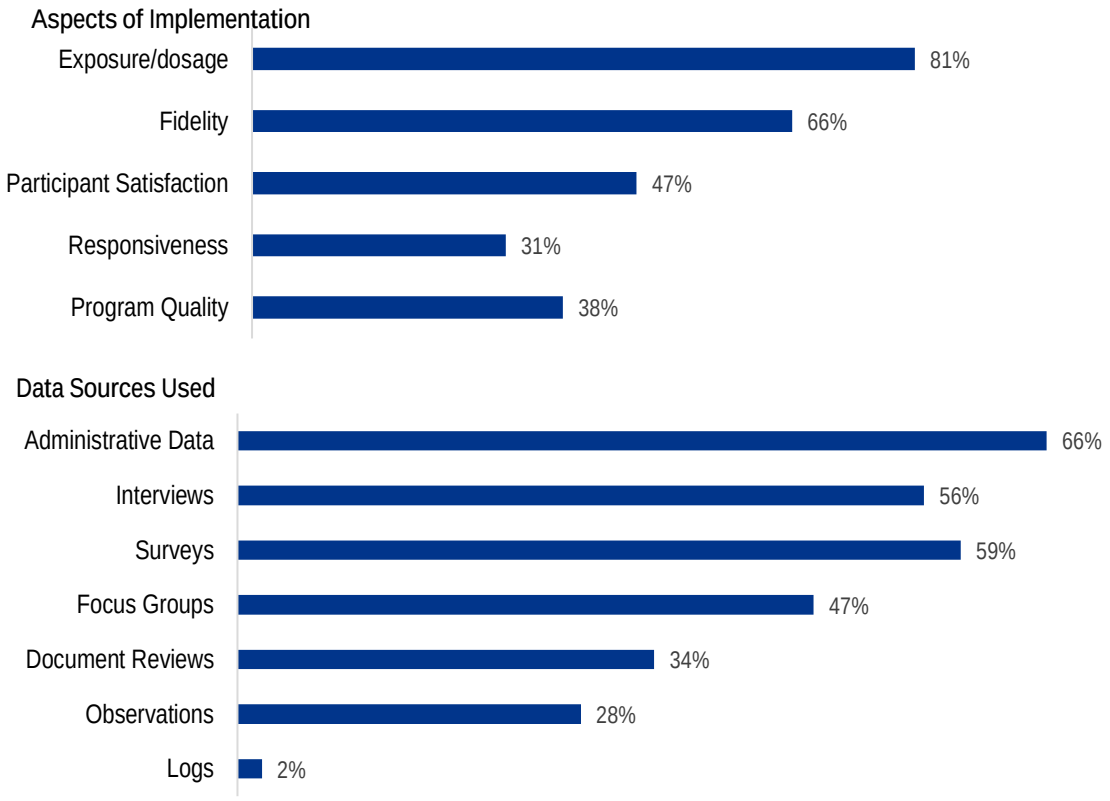
4.1 Implementation Evaluation Focus and Data Sources

Implementation evaluation assesses how well a program realizes what it sets out to put into practice. Rather than concentrating on outcomes or impact, implementation evaluations examine the process by which a program provides services. Implementation studies are not only important in their own right, but provide information essential for any impact/outcome study. For impact evaluations, Implementation data helps explore whether the intervention was implemented with fidelity to the program design and how it achieved (or did not achieve) its impact. This exploration may support or challenge claims of causality, examine the relationship between implementation and intervention effects, and identify challenges to future implementation and scale-up.

As noted in Exhibit 5, 32 reports provided evidence about program implementation, either as stand-alone implementation reports, or as joint implementation and impact/outcome reports. Exhibit 6 summarizes the aspects of implementation, as classified by Dane and Schneider (1998), that were examined in these reports. The analysis suggests that most reports addressed the exposure and fidelity aspects of implementation; half addressed participant satisfaction; and a third addressed program quality or responsiveness.

In terms of data sources, more than half the reports relied on administrative data, surveys, and interviews or focus groups. It is not surprising that observations and logs were used less frequently, because they require more resources and impose more burden on program staff or evaluators than other data collection alternatives.

Exhibit 6. *Implementation Evaluation Approaches (N=32)*



Note: Percentages do not add up to 100 because an evaluation can include multiple aspects and data sources.

The following examples illustrate how grantees and subgrantees combined multiple aspects of implementation and multiple data sources into a single implementation evaluation.

- **Youth development**. Pre-School-U's evaluation focused on exposure, fidelity, participant satisfaction, and program quality. Data sources included administrative data, interviews, observations, and logs.
- **Economic opportunity.** The evaluation of the Transitional Employment Services Model measured exposure, fidelity, and program quality, and drew information from administrative data and focus groups.
- **Healthy futures**. The evaluation of the Open Arms Children's Health program focused on exposure and participant satisfaction, with data from administrative records, surveys, and focus groups.

4.2 Impact/Outcome Evaluation

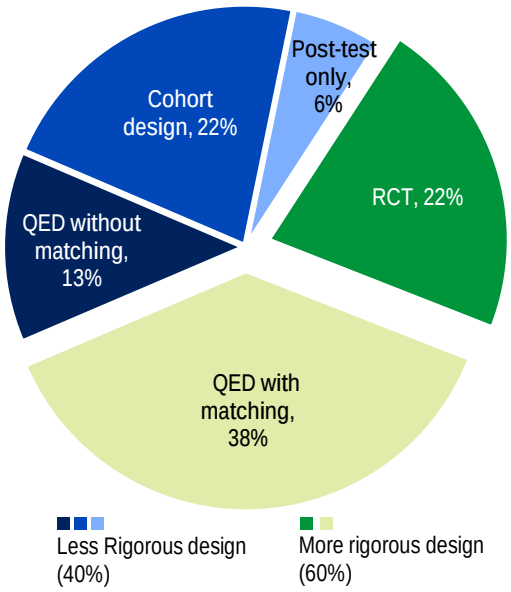
This section examines methodological aspects related to the impact/outcome evaluations, including research design, sample size and power, outcome measures, and analytic approaches to causal inference.

4.2.1 Research design

The quality of an impact study hinges on its research design, particularly on whether a comparison group is created through either random or non-random allocation. However, developing rigorous evaluations that lead to credible evidence of what works can be both costly and difficult. A CNCS report pointed out that, "Too often, effective nonprofits lack the expertise, resources, or infrastructure to evaluate their efforts, demonstrate impact, and take their programs from 'promising' to 'proven'" (Zandniapour and Vicinanza, 2013). For example, that report found that the SIF-funded RCTs cost an average of $437,110 per year—more than the entire annual subgrant for many SIF subgrantees. Projects that used community-based interventions, common in the healthy futures priority area, faced especially great challenges in aspects such as participant recruitment/retention, data collection, and making attributions, because participants may only receive one service rather than exposure to a more intensive or longer-term intervention.

As shown in Exhibit 7, of the 32 reports focusing on impacts/outcomes, 19 (60%) employed more rigorous designs such as RCT or QED design with matching.[4] Several additional reports indicated that the evaluation teams had originally proposed more rigorous designs, but could not implement them because of insufficient or reduced funding, difficulty collecting the outcome data, or problems in participant recruitment or retention—all common challenges in evaluation. To put these statistics in context, of the 72 SEPs approved by CNCS by 2014 (CNCS, 2015), 19% proposed to use non-experimental design, 47% QED, and 33% RCT.

Exhibit 7. *Impact/Outcome Evaluation Design (N=32)*



Note: Percentages may not add up to 100% due to rounding.

[4] For Exhibit 6, we classified a study according to its most rigorous design, if multiple designs were used, in order to make the categories mutually exclusive.

The SIF program began requiring grantees and subgrantees to design and implement impact evaluations that target moderate or strong levels of evidence with the 2011 cohort. Of the reports included in this synthesis, the following programs used the more rigorous RCT or QED with matching research designs.

- **RCT:** BELL Middle School Model, Reading Partners, Financial Opportunity Centers, SEED, SaveUSA, WorkAdvance, and Pathway to Literacy.
- **QEDs with matching** (most using propensity score matching): UC Degrees Gen-1 Project, Bridging the Gap, Resilient Children Project, Workforce Partnership Programs, Social Enterprises, Soccer for Success, Communities in Schools, Jeffco Summer of Early Literacy, CAS-Carrera, Financial Opportunity Centers, and Inkster Family Literacy Movement.

4.2.2 Sample and power

One of the considerations in impact evaluation is statistical power. A power analysis is a calculation that estimates, given a specific sample size and analysis approach, how likely it is that a program effect of a given size will be found to be statistically significant. Adequate power is essential for statistical validity because insufficient power may lead to an incorrect conclusion that the intervention does not have a significant impact, when it actually does. Although lack of power may not seriously threaten the internal validity of the findings, it will limit their generalizability.

Of 32 reports focusing on program impact/outcomes, the sample sizes and statistical power vary considerably. Even within one study, the samples may vary depending on the outcome measures. Half of the 32 impact/outcome evaluations appeared to be adequately powered in at least one outcome, while the remainder were under-powered. Below are a few examples of studies demonstrating adequate power.

- **Individual-level RCT.** Reading Partners (youth development): The final sample for this RCT study included 579 treatment students and 567 control students. The unit of assignment is individual students.
- **Individual-level QED.** Workforce Partnership Programs (economic opportunity): The final samples for this QED study using propensity score matching are separated into three sectors: health care (992 treatment, 46,701 comparison); advanced manufacturing (682 treatment, 42,293 comparison); and construction (379 treatment, 36,859 comparison). The propensity score matching allows multiple comparison subjects to be matched to one treatment subject.
- **Cluster-level QED.** Soccer for Success (healthy futures): For this QED study matched by sites, there are 712 treatment students from 16 sites and 522 comparison students from 14 sites. The unit of assignment is site.

4.2.3 Outcome measures

Ensuring that the chosen outcome measures are reliable, valid, and appropriate for a study is key to reducing threats to internal validity. The outcome measures in the SIF impact evaluations were closely aligned with the outcomes targeted by the interventions, and many of the studies employed valid and reliable measures to assess implementation and impact/outcomes.

Some of the outcome measures may be characterized as short-term or intermediate, others, as long-term. Many evaluations used a logic model to distinguish among these outcomes and conceptualize the relationships among them. Some used scales/instruments that are well-established in their fields, which helps address validity issues and allows comparison/integration with related research and evaluation literature. The examples below illustrate the types of outcomes and specific outcome measures examined in impact evaluations within each priority area.

- **Youth development.** Youth development outcome measures included student achievement or readiness, and student engagement or academic behavior/attainment measures such as attendance, course enrollment or completion, retention, and graduation. For example, Parents as Teachers (PAT) and Home Instruction for Parents of Preschool Youngsters (HIPPY) measured school readiness for preschoolers using the Ages and Stages Questionnaire communication subscale and MacArthur Communicative Development Inventories, both of which are validated instruments.
- **Economic opportunity**. Economic opportunity impact evaluations examined such outcomes as employment, job retention, earnings, job readiness, skill certifications, and life stability. For example, Social Enterprises employed two types of outcomes: 1) economic self-sufficiency (measured by employment and income) and 2) life stability (measured by housing situation, recidivism, and health outcomes).
- **Healthy futures.** Healthy futures impact evaluations tended to measure outcomes such as health outcomes, health behaviors, knowledge, self-efficacy, health care utilization, cost, and patient and stakeholder satisfaction. For example, the Get Healthy Access Program focused on health outcomes (e.g., blood sugar, blood pressure, body mass index, and waist circumference) and health-related behaviors (e.g., smoking, physical activity, and consumption of fruits and vegetables).

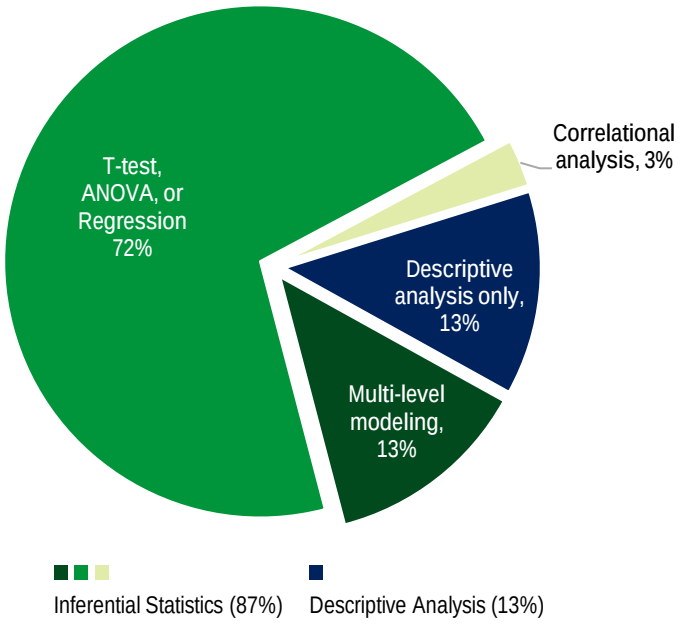
4.2.4 Statistical analysis for supporting causal inferences

Appropriate statistical analysis techniques ensure the rigor of evidence from the evaluation. The statistical analyses supporting the causal and correlational findings are often aligned with the study designs.

Descriptive analysis alone, while essential and informative, cannot establish causal or correlational inferences. Even among inferential statistics, some are more appropriate and rigorous than others, given the circumstances. For example, in an RCT, where the unit of assignment is a cluster (e.g., a classroom or school), a multi-level modeling approach is more appropriate than a single-level analysis, because it corrects for the cluster effect. In a group comparison, a multiple regression sometimes is a better approach than t-test or ANOVA, as it accounts for the other confounding variables that may be overlooked in estimates by t-test or ANOVA.

Of the 32 projects with impact/outcome focus, 87% used inferential statistics such as multi-level modeling, regression, t-test, ANOVA, or correlational analysis to support causal claims (Exhibit 8). We characterized a study by the most rigorous analysis methods, if multiple methods were included, in order to make the categories mutually exclusive.

Exhibit 8. *Analysis Approach to Causal Inferences (N=32)*

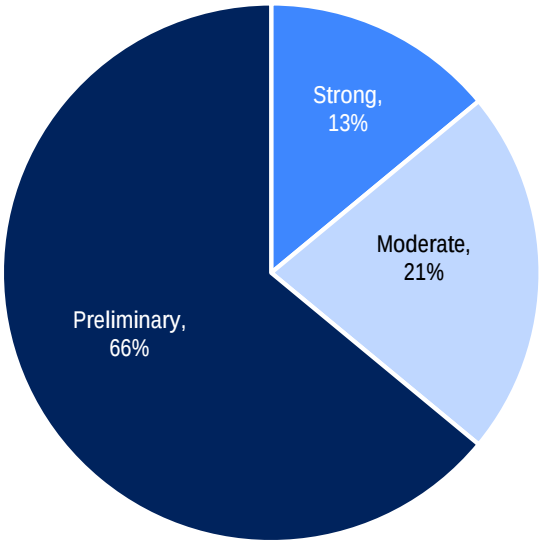


4.3 Level of Evidence

In order to support evaluation design and subsequent technical assistance, CNCS developed criteria to distinguish three tiers or levels of evidence: preliminary, moderate, and strong. Although other aspects of the evaluation—such as measures, sample size and retention, and data analysis—contribute to the rating, the level of evidence is primarily determined by evaluation design and by the attention the program's manner of implementation gives to threats to internal and external validity.

Exhibit 9 shows that, of the 38 evaluation reports reviewed, 5 were categorized as having achieved a strong level of evidence (Reading Partners, Soccer for Success, Communities in Schools, SaveUSA, and WorkAdvance), and 8 were rated as having achieved a moderate level of evidence (BELL Middle School Model, Workforce Partnership Program, Social Enterprises, SEED, Financial Opportunity Centers, Jeffco Summer of Early Literacy, Promotor Pathway Program, and Pathway to Literacy). The remaining 25 reports (including 6 with implementation-only results) were classified as falling within the preliminary level of evidence.

Exhibit 9. *Level of Evidence (N=38)*



5. Evaluation Findings

Evidence, especially from evaluations that yield a strong or moderate level of evidence, is likely to become one of the SIF's enduring legacies, because sound evidence will inform future innovation and scale-up by other nonprofits across the nation. Strong evaluations offer a route through which SIF/CNCS will contribute to the body of knowledge in highly important substantive areas. This section summarizes the substantive findings from these 38 reports, as well as information about implementation, impact/outcome, and cost. It also presents effect size estimates across programs overall and by subgroups.

5.1 Implementation Findings

The evaluation findings show that most SIF-funded programs were implemented appropriately with respect to the key aspects of implementation summarized in Exhibit 5. Of the 32 reports focused on implementation, most (88%) reported evidence of a high or adequate degree of implementation in terms of exposure/dosage, fidelity, and/or participant satisfaction. Below are three examples of projects with findings about implementation.

- **Youth development**. For Gateway to College, students consistently expressed strong appreciation for the program and valued the mature learning environment it offered. The program succeeded in maintaining a model that has a shared culture of student-centered support and a solution-focused pedagogy. The sites implemented the core model as designed. However, although flexibility in implementation allowed local programs to be responsive to student needs and build strong relationships with postsecondary host institutions, it created challenges in maintaining fidelity.
- **Economic opportunity**. The report for the Financial Opportunity Centers (FOCs) found that 66 % of the participants received assistance from the FOC counselors in at least one of the three core service areas during the two years after program entry. Thirty-five percent received assistance from the counselors in all three areas. Participation rates varied considerably across the five sites; participants who had at least a high school diploma or GED were significantly more likely to participate in the intended services than those who did not have a degree. Participants who were

ages 25 and older were more likely than those ages 18 to 24 to obtain assistance from any of the counselors and from the counselors in all three core service areas.

- **Healthy futures**. The St. Elizabeth Telepsychiatry program met the implementation goals, with 146 staff trained and reporting that they felt competent to provide services; 1,120 eligible patients who consented to receiving services; and 100% of patients in telepsychiatry assessment who received a plan of care, a referral for outside services, and/or follow-up care.

Some projects were not implemented according to plan. Below are examples that illustrate some of the difficulties encountered.

- **Participant recruitment.** For the Pre-school-U program, none of the sites reached the targeted recruitment and retention goals. The program did not seem to reach its target population either, because most participants were already competent in parenting and quickly lost interest after discovering that the program was not equipped to help them improve their skills.
- **Data collection.** For St. Joseph Health System, the evaluator concluded that the data were insufficient to determine implementation fidelity, because one site's data had not been collected at the time of the report.

5.2 Impact/Outcome Findings

Evaluation findings show that most SIF projects had positive effects on participants. Of the 32 reports concentrating on impact and/or outcomes, 94% found some positive effects. Exhibit 10 further reveals that 12 (38%) found positive effects on all outcomes, 18 (56%) produced mixed but mostly positive results, and 2 (6%) did not find any effects. The subset of evaluations with a strong or moderate level of evidence followed a similar pattern. Of the 13 evaluations rated as providing a strong or moderate level of evidence, 4 found positive effects on all outcomes, 8 yielded positive results on some but not other outcomes, and 1 concluded with no effect.

Exhibit 10. *Impact/Outcome Findings*

Impact/outcome findings	All evaluations (N=32)		Evaluations with strong or moderate evidence (N=13)	
	n	%	n	%
Some Positive Effect	30	94%	12	92%
Positive effect on all outcomes	12	38%	4	31%
Positive effect on some outcomes but not others	18	56%	8	62%
No Effect	2	6%	1	8%

Note: Totals do not total 100% due to rounding.

Listed below are illustrative examples of impact/outcome findings for three evaluations.

- **Positive effect on all outcomes**. The impact report from Early Childhood Science Inquiry Training for Educators (youth development) found the following: 1) treatment students scored higher than comparison students with no documented preschool attendance; 2) treatment teachers reported increased comfort and competency in the preparation of science and math lessons; and 3) parents reported greater comfort in having conversations with children about science, as well as increased awareness and ability to support science learning.
- **Positive effect on some outcomes**: Evaluation findings for the Meade Activity Center (healthy futures) suggested 1) statistically significant and positive impact on physical activity self-efficacy, physical activity behavior, sedentary behaviors, cardiorespiratory fitness level, and body mass index; 2) no impact on healthy eating behaviors.

- **No effect**. The impact evaluation of the Community Based Delivery Model/Virtual Care program (healthy futures) revealed no apparent changes in utilization of primary and specialty care, and no perceived improvements in health or self-efficacy.

When interpreting impact/outcome findings, it is important to note the associated level of evidence. For example, outcome findings from a report at the preliminary level of evidence should be interpreted as descriptive rather than causal.

5.3 Magnitude of Program Effect

Although the impact findings summarized in Section 5.2 are informative, it is difficult to derive from them a concrete sense of the magnitude of program effects across SIF-funded interventions. This section addresses that question by calculating effect size using statistics provided in the evaluation reports. By providing findings based on statistical calculation of effect sizes, this section of this report differs from most other sections, which are primarily based on thematic analysis.

Effect sizes are often used to measure the magnitude or practical significance of program impact. Effect size represents a standardized mean difference between treatment and comparison groups. Cohen (1988) defined effect sizes as "small, d = 0.2," "medium, d = 0.5," and "large, d = 0.8". However, Lipsey and Wilson's (2001) more recent compendium of meta-analyses concluded that psychological, educational, and behavioral treatment effects of modest values—even d=0.1 to 0.2—should not be interpreted as trivial.

Twenty-eight out of the 32 impact/outcome evaluations provide sufficient data to allow for calculation of effect size for one or more outcomes of interest. First, all evaluation reports were coded using a structured protocol. The protocol has four nesting levels (i.e., report, study, sample, and outcome measure). All of the outcomes were set in the same direction. That is, undesirable outcomes—such as increases in violence and delinquent behavior, unhealthy behavior, school drop-outs, housing instability—were flipped in direction during coding. Similarly, we flipped the signs of estimates for studies where higher numeric values represent undesirable results in measurement scales. Since many of the outcomes reported are intervention-specific, they were grouped into broader outcome domains for the purposes of this analysis. For example, outcomes such as reading achievement and math achievement, assessed by standardized tests, as well as GPA, were all categorized within the domain of "achievement and knowledge."

The analysis used the following domains, created based on the outcomes described in the evaluation reports.

1. Education/training participation (e.g., training, school, college enrollment)
2. Education/training completion (e.g., training, school, college completion, graduation, drop out)
3. Skill/knowledge application (e.g., use skills learned)
4. Achievement and knowledge (e.g., earned credit hours, test proficiency, GPA)
5. School readiness and aspiration (e.g., college intent and aspiration, kindergarten/college readiness)
6. Employment (e.g., employment attained)
7. Income
8. Job retention (e.g., teacher turnover, job retention)
9. Financial behavior (e.g., credit rating, debt, liquid asset, use of budget, having bank account)
10. Health behavior (e.g., vegetable consumption, physical activity, smoking, substance abuse, alcohol use, birth control, dental exam, eye exam, health checkup, pregnancy, condom use)

11. Health outcomes (e.g., BMI score, cardio test score, diabetes, depression, health score, hypertension, length of hospital stay, readmission, treatment acceptance, waist circumference, ER use)
12. Violence and delinquent behavior (e.g., crime, carry weapon, violence measure, arrest, physical fight)
13. Housing stability (e.g., sleep in shelter, halfway home, outside, permanent housing, transitional housing)

Then, effect size estimates were calculated using Comprehensive Meta-Analysis (CMA) software (Borenstein, Hedges, Higgins, and Rothstein, 2005). In addition to producing effect size estimates based on reported statistics, the software has a built-in weighting procedure that accounts for the sample sizes. Effect sizes from large samples are weighted more than those from small samples. However, the weighting for the combined effect estimates does not account for the study's level of evidence. The level of evidence was examined as a subgroup analysis.

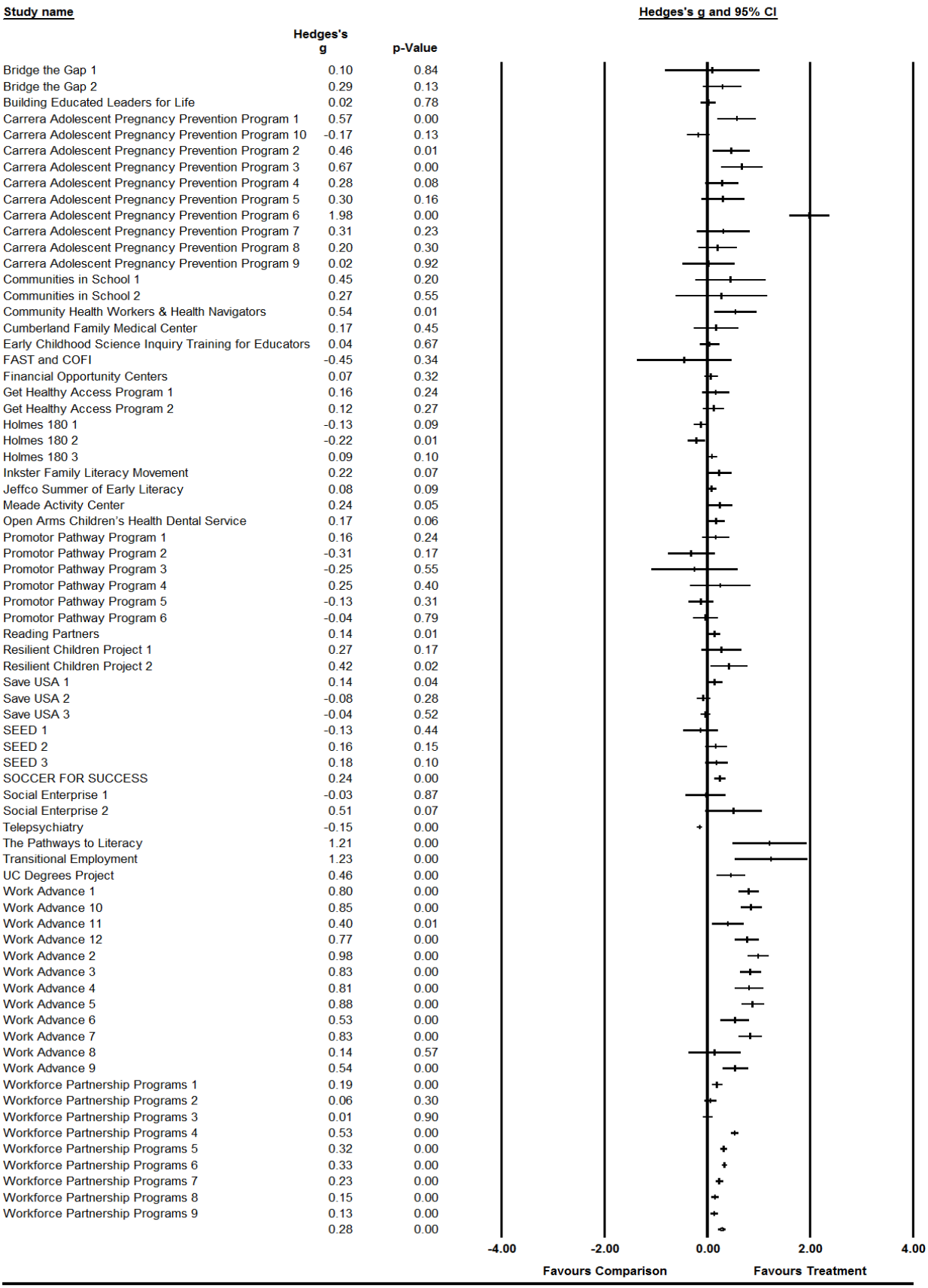
An impact evaluation of an intervention may consist of multiple impact studies on different outcomes from different samples. Each outcome from an independent sample within an evaluation was treated as a separate impact study. The 28 project reports examined in the meta-analysis contain findings from 72 independent samples with associated outcome domains based on results from 275 effect size estimates extracted from the evaluation reports.

Finally, the findings were reported not only by combined effect across all estimates and programs, but also by subgroups, including sectors, outcome domain, and level of evidence. Detailed methodological approaches are presented in Appendix A.

Combined effect. We conducted meta-analyses on the datasets of effect sizes to examine the overall effectiveness of the SIF across studies. Based on results from 72 effect size estimates of the outcome domains, Exhibit 11 shows that the average effect size was $g = 0.28$, and was statistically significantly different from 0. The effect size of 0.28 can also be interpreted as an 11 percentile gain by the treatment over comparison group (Marzano Research Laboratory, 2011). The overall impact is sizeable and meaningful.

Exhibit 11 lists each study. Some reports contain one study and other reports, multiple studies, as they examine multiple outcome domains and derive findings from different samples. For each study, two statistics were provided: Hedge's g measures the effect size (or the magnitude/practical significance of the effect); and p-value measures the statistical significance of the effect. The exhibit also includes a graphic display for each estimate showing the mean and the confidence interval.

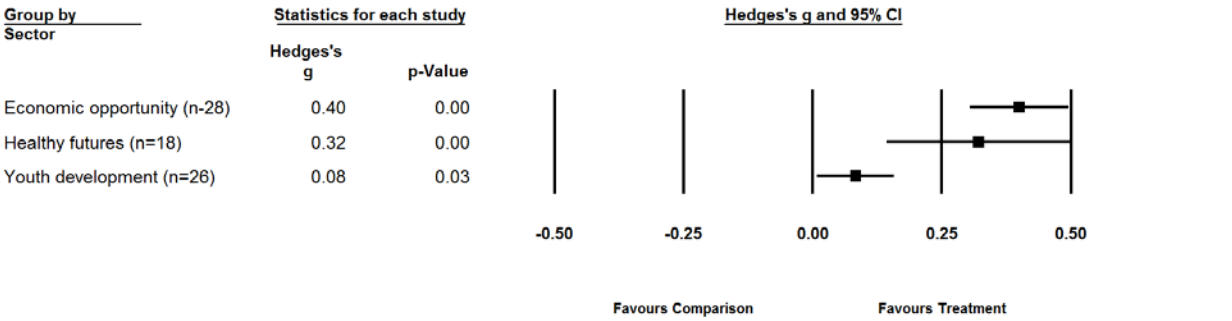
Exhibit 11. *Combined effect across programs*



Study name	Hedges's g	p-Value
Bridge the Gap 1	0.10	0.84
Bridge the Gap 2	0.29	0.13
Building Educated Leaders for Life	0.02	0.78
Carrera Adolescent Pregnancy Prevention Program 1	0.57	0.00
Carrera Adolescent Pregnancy Prevention Program 10	-0.17	0.13
Carrera Adolescent Pregnancy Prevention Program 2	0.46	0.01
Carrera Adolescent Pregnancy Prevention Program 3	0.67	0.00
Carrera Adolescent Pregnancy Prevention Program 4	0.28	0.08
Carrera Adolescent Pregnancy Prevention Program 5	0.30	0.16
Carrera Adolescent Pregnancy Prevention Program 6	1.98	0.00
Carrera Adolescent Pregnancy Prevention Program 7	0.31	0.23
Carrera Adolescent Pregnancy Prevention Program 8	0.20	0.30
Carrera Adolescent Pregnancy Prevention Program 9	0.02	0.92
Communities in School 1	0.45	0.20
Communities in School 2	0.27	0.55
Community Health Workers & Health Navigators	0.54	0.01
Cumberland Family Medical Center	0.17	0.45
Early Childhood Science Inquiry Training for Educators	0.04	0.67
FAST and COFI	-0.45	0.34
Financial Opportunity Centers	0.07	0.32
Get Healthy Access Program 1	0.16	0.24
Get Healthy Access Program 2	0.12	0.27
Holmes 180 1	-0.13	0.09
Holmes 180 2	-0.22	0.01
Holmes 180 3	0.09	0.10
Inkster Family Literacy Movement	0.22	0.07
Jeffco Summer of Early Literacy	0.08	0.09
Meade Activity Center	0.24	0.05
Open Arms Children's Health Dental Service	0.17	0.06
Promotor Pathway Program 1	0.16	0.24
Promotor Pathway Program 2	-0.31	0.17
Promotor Pathway Program 3	-0.25	0.55
Promotor Pathway Program 4	0.25	0.40
Promotor Pathway Program 5	-0.13	0.31
Promotor Pathway Program 6	-0.04	0.79
Reading Partners	0.14	0.01
Resilient Children Project 1	0.27	0.17
Resilient Children Project 2	0.42	0.02
Save USA 1	0.14	0.04
Save USA 2	-0.08	0.28
Save USA 3	-0.04	0.52
SEED 1	-0.13	0.44
SEED 2	0.16	0.15
SEED 3	0.18	0.10
SOCCER FOR SUCCESS	0.24	0.00
Social Enterprise 1	-0.03	0.87
Social Enterprise 2	0.51	0.07
Telepsychiatry	-0.15	0.00
The Pathways to Literacy	1.21	0.00
Transitional Employment	1.23	0.00
UC Degrees Project	0.46	0.00
Work Advance 1	0.80	0.00
Work Advance 10	0.85	0.00
Work Advance 11	0.40	0.01
Work Advance 12	0.77	0.00
Work Advance 2	0.98	0.00
Work Advance 3	0.83	0.00
Work Advance 4	0.81	0.00
Work Advance 5	0.88	0.00
Work Advance 6	0.53	0.00
Work Advance 7	0.83	0.00
Work Advance 8	0.14	0.57
Work Advance 9	0.54	0.00
Workforce Partnership Programs 1	0.19	0.00
Workforce Partnership Programs 2	0.06	0.30
Workforce Partnership Programs 3	0.01	0.90
Workforce Partnership Programs 4	0.53	0.00
Workforce Partnership Programs 5	0.32	0.00
Workforce Partnership Programs 6	0.33	0.00
Workforce Partnership Programs 7	0.23	0.00
Workforce Partnership Programs 8	0.15	0.00
Workforce Partnership Programs 9	0.13	0.00
	0.28	0.00

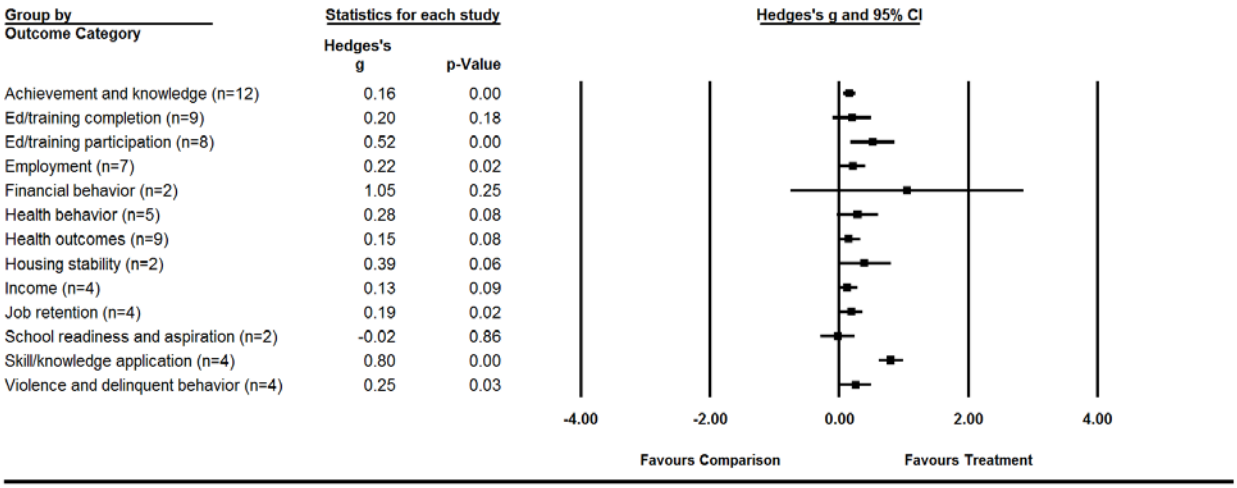
Effect by sector. Next, we looked at effect sizes classified by the intervention sectors (Exhibit 12). The average effect sizes were 0.08 for youth development, 0.40 for economic opportunity, and 0.32 for healthy futures, all of which were statistically significant. All three sectors have relatively large number of studies to support these estimates. The effect sizes for economic opportunity and healthy futures are considered to be medium. The effect size for youth development interventions is small, largely driven by results from interventions such as Holmes 180, FAST, and CORT that pulled the average effect size downward.

Exhibit 12. *Effect by Sector*



Effect by outcome domain. Next, we looked at effect sizes organized by outcome domains. Of the 13 outcome domains examined, some outcomes had only 2 data sources while others such as achievement and knowledge drew on results from as many 12 studies. Exhibit 13 shows that the estimated effect sizes ranged from -0.02 on school readiness to 1.05 on financial behavior. Results from 10 out of 13 outcome domains were statistically significant at 0.10 level. The non-significant findings were partly due to small sample sizes. For example, outcome domains such as financial behavior, housing stability, school readiness, and aspiration only have 2 studies.

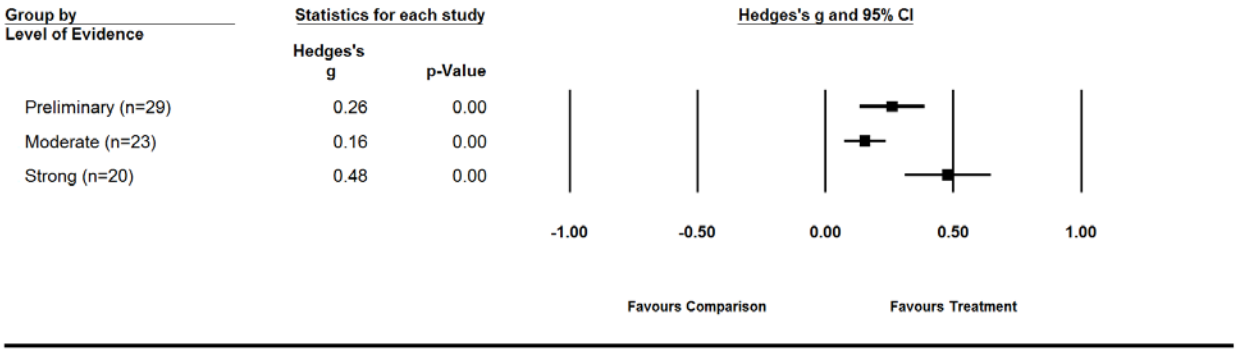
Exhibit 13. *Effect by Outcome Domain*



Effect by level of evidence. Finally, we examined the association between the level of evidence and the impact estimates. Exhibit 14 shows that the average effect sizes from studies with preliminary evidence is 0.26, moderate evidence 0.16, and strong evidence 0.48, all of which are statistically significant. All three groups have large numbers of studies to support these estimates. The fact that effect sizes from

studies with strong evidence are larger than those with preliminary or moderate evidence, is not consistent with other meta-analyses, which suggest that studies with weaker designs tend to produce larger effect sizes than those with stronger designs (Chueng and Slavin, 2016; Zhang, 2013).

Exhibit 14. *Effect by Level of Evidence*



5.4 Cost Findings

At a time of limited resources and competing priorities, findings about program cost are important for replication and scale-up, particularly for interventions yielding positive impact with strong or moderate levels of evidence. Although cost analysis was not required as part of the SIF evaluation, it was included in 6 of the 38 reports (16%). Each of these programs was found to offer cost-effective solutions to addressing the target areas:

- **Reading Partners** (youth development). The program is a low-cost option for under-resourced schools because in-kind contributions from community volunteers pay most of the costs. On average, schools bear about 20% ($710 per program student) of the total cost of the resources required, more than half of which are in-kind contributions from the school of space and staff time. The evaluation, supported by a strong level of evidence, found positive impact on all outcomes.
- **Jobs Plus** (economic opportunity). The cost-per-resident-household of operating the program at the targeted housing units in year 3 was $672 in New York and $503 in Texas. Financial resources to support the program staff's implementation and engagement among clients and community members was sizable, i.e., over 70% of funds. The evaluation, supported by a preliminary level of evidence, focused only on implementation.
- **Social Enterprises (SE)** (economic opportunity). The return on investment (ROI) for the SE program was at least as large, if not larger, than estimated ROI from similar programs. For every dollar SE spent, the SE returned $2.23 (outcome study) or $1.34 (impact study) in total benefits. The evaluation, supported by a moderate level of evidence, found positive impact on most but not all outcomes.
- **Community Based Delivery Model/Virtual Care** (healthy futures)**.** For specialty care, patients averaged a savings of $38 per consultation by utilizing telemedicine. The evaluation, supported by a preliminary level of evidence, did not find any impact on health outcomes.
- **Project Rise** (economic opportunity). An analysis of costs incurred during the operation of cohorts two through five found that Project Rise, on average, cost $6,636 per enrollee to operate. Nearly two-thirds of this cost was associated with staff salaries and fringe benefits.

- **WorkAdvance** (economic opportunity). The evaluation considered gross and net costs of operating the program across seven dimensions. On average, the gross cost ranged from $5,000 to $6,000 per participant at all four sites. Most of these costs were allocated to providing pre-employment and occupational skills training services. These differences are attributable to differences in the costs of the occupational skills training programs at community colleges, for-profit providers, and four-year colleges in which control group members enrolled.

6. Reflections from Evaluators

In addition to substantive findings, the evaluation reports examined for this meta-analysis reflected on various issues, including challenges, limitations, lessons learned, and potential contributions to the evidence base. We briefly discuss these reflections here, because they provide valuable additional insights for future evaluations.

6.1 Challenges in Program Implementation and Evaluation

About 40 % of the reports discussed challenges encountered during program implementation and evaluation, and in some cases, multiple challenges. Challenges related to program implementation included delays in schedule due to SEP approval and activities such as receiving program materials, difficulties with participant recruitment and retention, concerns about implementation fidelity, and other issues such as transportation to sites, staffing, and grant management (such as complexity in invoicing). Challenges related to evaluation included issues with data collection resulting from limited funding, difficulty achieving desired response rates, challenges in managing tracking systems, and other unexpected roadblocks, including delays in SEP approval, evaluator turnover, and use of an inappropriate evaluation instrument. Below are examples.

- **Participant retention and recruitment**. Gateway to College's (youth development) biggest challenge was retaining students in the program during the initial Foundation term. Fewer than half the students passed all their Foundation courses and successfully transitioned to mainstream community college. The most serious stumbling block was English Language Arts. Retention was affected by many pressures and problems that existed prior to enrolling in the program, such as health, family issues, and work conflicts.
- **Data collection**. Cumberland Family Medical Center (healthy futures) experienced significant challenges with the electronic health record (EHR) system and the process of implementing a research program in a rural health care delivery system. The evaluation team could not obtain the information needed for the evaluation from the EHR—a patient-focused system not designed to support research.
- **Limited funding**. King's Daughter Medical Center (healthy futures) encountered funding limitations, which curtailed data collection and prevented the evaluators from completing the evaluation as originally proposed in the SEP.
- **Early termination.** For the Inkster Family Literacy Movement (youth development), the project ended earlier than planned, and data collection only took place over one month. No adequate comparison group was formed, and the sample sizes for the survey and assessment were very small.
- **Change of intervention in response to program, SIF, or another external requirement.** For Families and Schools Together (youth development), the program requirement that participants live within a certain zip code reduced the number of potential participants significantly, given the fluid boundaries that exist across River Rouge and its neighboring communities. In addition, the original

design was modified to meet SIF guidelines for the moderate level of evidence. This mid-course change, which occurred during year two, caused a shift in the theory of change as well as staffing and approach. As a result of the decision to eliminate the family literacy program, the focus of FAST recruitment changed to Head Start teachers, which proved to be an effective strategy.

6.2 Limitations of Evaluations

About 40 % of the reports explicitly discussed limitations, most of which concerned evaluation approaches. These limitations included lack of counterfactual, data collection problems with measures, participant recruitment and retention, data quality issues such as missing data and low response rates, lack of statistical power, inability to address important questions, conducting the evaluation at sites implementing the program for the first time, and significant program changes occurring after the evaluation data collection was completed. Some examples are:

- **Lack of counterfactual**. Bridging the Gap (youth development) could not implement propensity score matching as proposed. There was selection bias between treatment and comparison groups as well as confounding factors in attribution.
- **Participant attrition and missing data**. Early Childhood Science Inquiry Training for Educators (youth development) found that uncertainty in the lives of those in the treatment group resulted in considerable attrition, reduction of sample size, and missing data.
- **Target populations**. Although the Center for Employment Opportunities (economic opportunity) has been shown to be most effective for those within three months of release from incarceration, the sites were not so successful at targeting those recently released or at highest risk. Unlike New York City, with its extensive public transportation system, the limited transit in several communities restricted their ability to recruit or retain members of the target population.
- **Program changes and statistical power**. Building Educated Leaders for Life (youth development) evolved significantly after the evaluation data collection, rendering the results less relevant. The study, underpowered from the outset, was also conducted in school districts implementing the program for the first time, raising questions about fidelity.
- **Program adaptation.** CAS-Carrera (healthy futures) adapted a successful after-school program to the in-school setting. Children enter the integrated school model by virtue of their grade level or administrator choice and do not choose membership, as was the case in the after-school settings. Additionally, the in-school model has a different student-to-facilitator ratio than the typical after-school CAS-Carrera program. Thus, the typical one-on-one relationships that exist in an after-school program are more challenging to develop in the in-school model.

6.3 Lessons Learned

More than half the reports specifically discussed lessons learned and made recommendations. These reflections were often based on perceived successes, setbacks, or ideas for future exploration.

- **Program-related lessons** included the need to:
    - Prioritize efforts to improve working relationships among stakeholders
    - Plan strategies to maximize recruitment and retention of participants
    - Seek to provide a longer duration of intervention, to allow sufficient time to achieve a measurable impact

- Keep the program small at start-up to allow for rapid adjustments and program improvements, and scale up once the initial complications are resolved
- Plan for a long start-up period with substantial technical assistance
- Consider cost implications explicitly when making program decisions
- Assess the infrastructure improvements needed for effective implementation and tackle them early in the program
- Find ways to expand successful programs to different populations, remaining alert to adjustments necessary to account for differences (e.g., age or cultural differences)
- Recognize the importance of program adaptation, flexibility, and individualization in scale up.

- **Evaluation-related lessons** include the need to:

  - Provide appropriate levels of funding for data collection
  - Obtain staff with the skills to handle large datasets
  - Plan for tracking longitudinal outcomes
  - Select sites experienced with the program to enhance fidelity
  - Collect detailed information about the control condition to allow for appropriate comparisons
  - Obtain methodological guidance early and devise sound data collection methods to ensure that the study is well-powered
  - Consider potential and stronger impact beyond the study timeframe and for higher need populations.

6.4 Contributions to the Evidence Base

About half the reports explicitly addressed how they contributed to the evidence base in the field. All the implementation reports discussed issues relating to the implementation, replication, or scale up of a particular intervention in a target area. Consistent with their levels of evidence, impact/outcome evaluations made very concrete contributions, such as the following:

- **Offering evidence for the effectiveness of a low-cost option**. The Reading Partners (youth development) report explores the low-cost option of using community volunteers to provide one-on-one tutoring. One-on-one tutoring by certified teachers has been demonstrated to be effective in improving the reading proficiency of struggling readers, but it is both time- and resource-intensive.
- **Providing empirical evidence to a field with little pre-existing research**. The Social Enterprises study (economic opportunity) expands the information available about the effectiveness and costs/benefits of social enterprises, about which little impact research exists.
- **Advancing the level of evidence**. The evaluation of the Promotor Pathway program (youth development) used more rigorous research approaches, advancing the evidence level of the Latin American Youth Center from an incoming level of preliminary to moderate.
- **Replicating a model in a new context.** The Hope Health Clinic (healthy futures) replicated the model developed by the Church Health Center in Memphis that had operated for 25 years but was never systematically evaluated. The model builds on the literature concerning the benefits of providing low-cost and accessible primary care to uninsured persons, and the use of lay patient navigators, community health workers, non-clinicians, and clinic volunteers to improve patient self-efficacy and access to services.

7. Conclusions

This meta-analysis of 38 evaluations of SIF-funded interventions offers important evidence about the implementation and impact of these programs as well as valuable insights about the evaluations. We conclude with five major conclusions.

The SIF catalyzes innovation in targeted sectors. Thirty-two percent of the reports were in the healthy futures issue area, 45% in youth development (including early childhood), and 24% in economic opportunity. Funded programs employed a variety of programmatic strategies for target populations, including summer programs, one-on-one tutoring, social enterprises, workforce partnerships, community-based health care, and telemedicine.

SIF evaluations employed rigorous methods to assess program implementation and impact. Thirty-two of the 38 reports evaluated implementation, and 32 evaluated outcomes or impacts (17 reported on both). Fifty-six percent of the 32 impact/outcome evaluations employed such rigorous designs as randomized controlled trial (RCT) or quasi-experimental (QED) with matching, which allows causal inferences. Half of the impact/outcome evaluations appeared to be adequately powered in all or at least one outcome. Most of these evaluations used valid and reliable outcome measures to capture the expected outcomes of interventions, and employed inferential statistics to support the findings. Of the 38 final evaluation reports, 5 were categorized as having produced a strong level of evidence, 8 as providing a moderate level, and 25 (including 6 with implementation only results) as providing a preliminary level.

Early findings show that SIF interventions generated meaningful and statistically significant impact, overall and across sectors. The evaluation findings show that most of the SIF-funded programs (87%), for which evaluation reports are available, were implemented with fidelity and had a positive impact on all or some target outcomes.

Meta-analysis of impact findings from 28 impact/outcome reports that provide sufficient statistics show statistically and practically significant effects. The average effect size was 0.28 and was statistically significant. This can be interpreted as 11 percentile gain by the treatment over comparison group. The average effect is the largest in economic opportunity intervention (0.40), followed by healthy futures (0.26) and youth development (0.08), all of which were statistically significant. We characterized outcomes evaluated by 38 reports into 13 broad outcome domains. Results from 10 out of 13 outcome domains were statistically significant at 0.10 level. The non-significant findings were partly due to small sample sizes. Finally, sizes from studies with preliminary evidence is 0.26, moderate evidence 0.16 and strong evidence 0.48, all of which were statistically significant.

SIF evaluations advanced knowledge about program implementation and evaluation in substantive areas. Between 40 to 50 % of the reports reflected on challenges, limitations to the findings, and lessons learned. For example, program-related lessons included suggested improvements to program elements and to working relationships among stakeholders, and other practices related to participant recruitment and retention, length of intervention, modest start-up, cost implications, infrastructure, and expansion to different populations. Evaluation-related lessons concerned adequate funding for data collection, sizeable administrative datasets, tracking longitudinal outcomes, and others, such as statistical power, detail about the control condition, and selection of inexperienced sites.

Meta-analysis provides a rigorous and efficient tool to synthesize findings across evaluations. The evaluations analyzed in this reports represent about a third of the evaluations by the SIF in the pipeline. The findings from this report are mostly from the 2010 cohort. SIF grantees and subgrantees in subsequent cohorts have received clearer expectations and guidance as well as stronger technical assistance for rigorous evaluations. It is reasonable to expect that evaluations from later cohorts will be

more rigorous and provide greater contribution to the knowledge base. It is also important to note that the SIF has a fully operational pipeline of on-going evaluations that will generate more results in the coming years to strengthen the evidence base in the nonprofit sector.

Findings from this report have important implications to the stakeholders.

- **For policymakers**, the statistically and practically significant results across evaluations validate the significant CNCS investment in evaluating different program models designed to improve a wide range of outcomes and nurture impact. As more evaluations are completed, policymakers should continue to ensure that the findings are synthesized and disseminated in order to develop the knowledge base and distill these lessons into accessible guidance for nonprofits seeking to implement evidence-based interventions.
- **For philanthropy and nonprofits.** The findings demonstrate the power of rigorous evaluations. Practitioners should recognize the importance of not only providing services to populations in need, but building evidence to inform replication and scale up for wider implementation. These evaluations provide valuable lessons for strengthening program implementation, including improving relationships among stakeholders, articulating programmatic strategies and theories of change, considering cost implications when making program decisions, and recognizing the importance of adaptation, flexibility, and individualization in scaling effort. They also highlight the importance of assisting nonprofit organizations to adopt the practices of promoting evidence-based programs and evaluations.

- **For evaluators.** SIF provides an opportunity and funding to test the efficacy of innovative programs that have at least preliminary evidence of effectiveness. The meta-analysis approach illustrates the contribution these individual evaluations can make to growing the evidence base. These evaluations also provide crucial lessons for strengthening evaluation, such as exploring the full range of design options in light of questions of interest, rigor, and feasibility; recognizing the importance of conducting an implementation study in addition to an impact study; and anticipating potential challenges and developing solutions by thinking through all aspects related to data collection, including data access, sampling, recruitment, and retention.

References

Borenstein, M., Hedges, L.V., Higgins, J., and Rothstein, H. (2005). *Comprehensive meta-analysis, version 2*. Englewood, NJ: Biostat.

Cheung, A., and Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5): 283-292.

Corporation for National and Community Service. (2013). *Getting results, transforming lives: the Social Innovation Fund, 2010-2012 investment report*. Washington, DC: Corporation for National and Community Service.

Corporation for National and Community Service. (n.d.) *Social Innovation Fund: Evaluation plan guidance-A step-by-step guide to designing a rigorous evaluation*. Washington, DC: Corporation for National and Community Service.

Corporation for National and Community Service, Office of Research and Evaluation. (2015). The Social Innovation Fund: Pioneering an Evidence-Based Investment Model (by Christopher Spera, Adrienne DiTommaso, Mary Hyde, and Lily Zandniapour). Washington, DC: Corporation for National and Community Service.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Cooper, H., Hedges, L.V., and Valentine, J.C. (Eds.) (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Dane, A.V., and Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1): 23-45.

Glass, G.V., McGaw, B., and Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Konstantopoulos, S., and Hedges, L.V. (2009). Analyzing effect sizes: Fixed-effects models. In H. Cooper, L.V. Hedges, and J. C. Valentine. (Eds.) *The handbook of research synthesis and meta-analysis,* 2nd ed. (pp. 279–294). New York: Russell Sage Foundation.

Lester, P. (2015). *Social innovation fund: early results are promising*. Social Innovation Research Center.

Lipsey, M.W., and Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Mazarno Research Laboratory. (2011). *Conversion of effect size to percentile gain.* Retrieved from http://soltreemrls3.s3-website-us-west-2.amazonaws.com/marzanoresearch.com/media/documents/reproducibles/highly_engaged/conversionofeffectsize.pdf

Steenbergen-Hu, S.S., and Cooper, H. (2012). A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. Unpublished manuscript.

Valentine, J., and Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality and intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods,* 13(2), 130–149.

Zandniapour, L., and Vicinanza, N. (2013) *Budgeting for rigorous evaluation: Insights from the Social Innovation Fund*. Washington, DC: Corporation for National and Community Service, Office of Research and Evaluation.

Zhang, X. (2015). *Meta-synthesis of evaluations across Social Innovation Fund (SIF): Interim report*. (Prepared under contract to Corporation for National and Community Service). Fairfax, VA: ICF International.

Zhang, X., Tran, H., Chen, E., Steketee, M., Hershey-Arista, M., Katterfeld, K., Ristow, L., and Frechtling, J. (2013). *Quality review and meta-analysis of impact studies for the National Science Foundation's Discovery Research K-12 (DR K-12) Program*. (Prepared under contract to the National Science Foundation). Rockville, MD: Westat.

Appendix A: Meta-Analysis Methodology

To estimate the effects of SIF-funded interventions, we used meta-analysis techniques. Impact and outcome evaluations were coded and analyzed to describe the methods used for study and calculate the effects of these projects.

Meta-analysis of program evaluations has taken on increasing policy importance. Compared to evaluations that use a singular design (e.g., randomized controlled trial) and rely on primary data collection, meta-analysis provides a more flexible and cost-effective, less burdensome approach. This is especially true for a complex program such as the SIF, where the heterogeneity of its projects makes it impossible and inappropriate to use a singular research design. It is also much more rigorous than the traditional literature review, which tends to focus on selective findings or synthesize findings from a large body of literature using bean-counting methods, enumerating the number of studies with positive, negative, or non-significant results.

In addition, meta-analysis has specific features well-suited for this evaluation. First, through a systematic research synthesis, the influence of context on the outcomes of policy, program, and practice evaluations can be examined by comparing studies varying in participant, setting, and treatment characteristics. Context effects can be examined in research reviews even though no single study contains all the variations. Second, multiple studies can also be grouped according to the strengths and weaknesses of their designs. If studies with different strengths and weaknesses lead to similar results, greater confidence can be placed in a review's conclusion than in the results of any single evaluation. If results are different, rival hypotheses can be precisely identified for testing through future study. Finally, by statistically combining the results of multiple studies, the general effect of a policy, program, or practice can be identified much more precisely than in a single investigation. In the remaining sections of this appendix, we provide a detailed discussion on several methodological issues related to the meta-analysis in the report.

Instrument and Coding

The protocol development was informed by other established rubrics used by the lead author (Zhang, 2013), and by Cooper (2010), Lipsey and Wilson (2001), and Steenbergen-Hu and Cooper (2012). The protocol has four nesting modules (i.e., report, study, sample, and outcome measure). For example, a project impact report may include multiple impact studies (e.g., impact from overall intervention and/or by elements of intervention). A study may report results separately for different samples (e.g., overall sample or by subgroups defined by gender, socioeconomic status, or other relevant characteristics such as location). A study may simultaneously examine multiple outcome measures (e.g., impact on student achievement, employment) for each independent sample. This meta-analysis includes only interim and long-term outcomes and excludes short-term outcomes such as satisfaction or participation (unless participation is the long-term outcome). All of the outcomes were set in the same direction. Undesirable outcomes—such as increases in school dropout, violence and delinquent behavior, unhealthy behavior (smoking, drug use), housing instability—were flipped in direction during coding. Similarly, the signs of estimates were flipped for studies where higher numeric values represent undesirable results in measurement scales. Finally, since many of the outcomes reported are intervention-specific, they were grouped into broader outcome domains for the purposes of this analysis. For example, outcomes such as reading achievement and math achievement, assessed by standardized tests, as well as GPA, were all categorized within the domain of "achievement and knowledge." The analysis used the following domains, created based on the outcomes described in the evaluation reports.

1. Education/training participation (e.g., training, school, college enrollment)
2. Education/training completion (e.g., training, school, college completion, graduation, drop out)
3. Skill/knowledge application (e.g., use skills learned)
4. Achievement and knowledge (e.g., earned credit hour, test proficiency, GPA)
5. School readiness and aspiration (e.g., college intent and aspiration, kindergarten/college readiness)
6. Employment (e.g., employment attained)
7. Income
8. Job retention (e.g., teacher turnover, job retention)
9. Financial behavior (e.g., credit rating, debt, liquid asset, use of budget, having bank account)
10. Health behavior (e.g., vegetable consumption, physical activity, smoking, substance abuse, alcohol use, birth control, dental exam, eye exam, health checkup, pregnancy, condom use)
11. Health outcomes (e.g., BMI score, cardio test score, diabetes, depression, health score, hypertension, length of hospital stay, readmission, treatment acceptance, waist circumference, ER use)
12. Violence and delinquent behavior (e.g., crime, carry weapon, violence measure, arrest, physical fight)
13. Housing stability (e.g., sleep in shelter, halfway home, outside, permanent housing, transitional housing)

All of the studies were coded by the lead author. The first two authors have participated in a training provided by Dr. Michael Borenstein, developer of the software Comprehensive Meta-Analysis (CMA) (Borenstein, Hedges, Higgins, and Rothstein, 2005), that addressed both conceptual issues related to meta-analysis as well as the use of the CMA software.

Analytic Approaches

This section provides detailed discussions and justifications of analytic approaches including effect size calculation, unit of analysis, and subgroup analysis.

Effect Size Calculation. Effect sizes are often used to measure the magnitude or practical significance of program impact. Cohen (1988) defined effect sizes as "small, d = 0.2," "medium, d = 0.5," and "large, d = 0.8". However, Lipsey and Wilson's (2001) more recent compendium of meta-analyses concluded that psychological, educational, and behavioral treatment effects of modest values—even d=0.1 to 0.2—should not be interpreted as trivial.

Hedges' *g*, a standardized mean difference between two groups, was used as the effect size index for this meta-analysis. The preference for Hedges' *g* over other standardized-difference indices, such as Cohen's *d* and Glass's *Δ*, is due to the fact that Hedges' *g* is corrected to reduce the bias that may arise when the sample size is small (i.e., $n < 40$; Glass, McGaw, and Smith, 1981).

Hedges' *g* was calculated by subtracting the mean of the comparison condition from that of the treatment condition and dividing the difference by the average of the two groups' standard deviations. A positive *g* indicates that the treatment has a positive effect on the outcome variable when compared to those in the comparison condition. In cases for which only inference test results were reported, but no means and standard deviations were available, *g* was estimated from the inferential statistics, such as *t*, *F*, or *p*-values (Lipsey and Wilson, 2001). For studies that did not report specific values of inferential statistics, we assumed a conservative value for effect size calculation. For example, if a study reported a statistically significant difference between the treatment and the comparison condition with $p < .01$, we assumed a *p*-value of 0.01 for effect size calculation.

The effect size calculation used a weighting procedure and random-effects models. A random-effects model assumes that the intervention has more than one true effect, and that the effect sizes included in a meta-analysis are drawn from a population of effects that can differ from each other. Fixed-effect models were not used because they assume that there is one true effect in all the studies included in a meta-analysis, and the average effect size will be an estimate of that value. A fixed-effect model is suited to drawing conditional inferences about the observed studies. However, it is less well suited to making generalizations to the population of studies from which the observed studies are a sample (Konstantopoulos and Hedges, 2009). In the case of SIF projects, the fixed-effect assumption is even more problematic because the interventions vary from one project to another.

Unit of Analysis. Meta-analysis requires using independent samples as the unit of analysis. Without using such a correction, data from studies using multiple outcomes from the same sample are treated as multiple studies. Each independent sample with its outcome domain is the equivalent of a separate research study. For example, if there are independent subgroups within a study, each subgroup contributes independent information. If there are effect sizes for more than one outcome within a study, based on the same participants, the information for the different effects is not independent, which needs to be taken into account in the analysis. The 28 project reports examined in the meta-analysis contain findings from 72 independent samples with associated outcome domains extracted from 275 effect size estimates from the report.

Moderator/Subgroup Analysis. The purpose of testing for moderators was to identify variables associated with certain features of the primary studies that might be significantly associated with the effectiveness of treatment (Cooper, Hedges, and Valentine, 2009). These moderators are related to either the intervention (e.g., sector focus) or design of the study (e.g., research design, level of evidence). In addition to reporting findings of the combined effect across all estimates and programs, we reported effect by subgroups, including sectors, outcome domains, and levels of evidence.

Appendix B: Programmatic Strategies from Reviewed Interventions

Program name	Programmatic strategies
Youth Development	
BELL Middle School Model	**Summer program** for middle school students to improve student academic achievement and engagement
Gateway to College	**Dual-enrollment program** for at-risk students to improve high school graduation and college readiness
Reading Partners	**One-on-one reading tutoring** to struggling readers by community volunteers
Early Childhood Science Inquiry Training for Educators (ECSITE)	**Pre-school curriculum and professional development**, which integrate science experiences across other curriculum domains through enriched environment, planned lessons, and incidental conversations with children
Holmes 180	**Comprehensive high school reform** designed to connect high school students with meaningful career opportunities through instructional and policy changes
UC Degrees Gen-1 Project and Higher Education Mentoring Initiative	**Residential program** providing low-income first-generation college students and foster youth with a variety of academic, supportive, and other services
Parents as Teachers (PAT), and Home Instruction for Parents of Preschool Youngsters (HIPPY)	**Home visiting programs** designed to increase the frequency with which parents engage in age-appropriate language and other activities with their children
Pre-School-U	**Program to train parents and caregivers** with knowledge and skills to best prepare their children for school, implemented in faith-based settings
Bridging the Gap	**Program providing work experience and professional development in job and life skills** to reduce dropout and improve adult employment for high school students through partnership with employers and community service providers
Resilient Children Project	**School readiness program** integrating educational and mental health services through early childhood mental health consultation services designed to increase teacher and caregiver competence
SEED Charter Schools	**Charter school** that offers a fully-integrated academic and boarding program to provide scheduled study time, constant access to positive role models, and life skills training
Communities in Schools	**Drop-out prevention program** targeting both school-wide intervention and individual student needs
Jeffco Summer of Early Literacy	**Voluntary summer program** that aims to improve the literacy of children who are entering kindergarten through third-grade at four low-income schools
Promotor Pathway Program	**Youth outreach program** that connects youth with a "promotor" who provides intensive case management, mentorship, and advocacy to help high-risk and disconnected youth overcome significant life obstacles
Pathway to Literacy (PTL)	**Combined activities** to improve school readiness of preschoolers through **home-based parent visitation** services using PTL curriculum; **play groups**; and **child development workshops**
Inkster Family Literacy Movement	**Combined activities to improve school readiness through community activities, awareness marketing campaigns, and community coalition partnerships** for preschoolers
Families and Schools Together (FAST); Community Organizing Around Family Issues (COFI)	FAST engages parents in **learning and support groups** to enhance family functioning and reduce parenting stresses; COFI uses parents' strengths and commitment to their children and neighborhood to help make positive changes in their own lives, families, and communities.
Economic Opportunity	
Center for Employment Opportunities (CEO)	**Transitional jobs program** for parolees
Workforce Partnership Programs	**Program establishing local workforce partnerships** to help low-income workers to obtain skills in needed areas

Program name	Programmatic strategies
Jobs-Plus	**Employment and training, financial incentives, and community support** to public housing beneficiaries
Social Enterprises	**Social enterprises**—businesses sell goods and services that the market demands in order to intentionally employ individuals who would otherwise have difficulty obtaining employment: providing employment, employment support while employed, life stability supports while employed, and post-employment support
Transitional Employment Services Model	**Employment and training** to allow hard-to-employ individuals to simultaneously access vocational skills training, work readiness training, and job seeking assistance.
Project Rise	**Combination of case management, GED instruction, work readiness training, and a paid 18-week internship** designed to enhance the success of at-risk, out-of-school young people by re-engaging them with education and the labor market
SaveUSA	**Tax-time savings program** which offers low and moderate-income families the opportunity to directly deposit all or a portion of their tax refund into a special savings account and pledge to save a specific amount for about a year
WorkAdvance	**Combination of pre-employment and career readiness services, occupational skills training, job development and placement services, post-employment retention and advance services**
Financial Opportunity Centers	**Combination of financial counseling, employment services, and assistance** in accessing other social welfare services with an aim to improve the overall financial well-being of low income families
Healthy Futures	
Cumberland Family Medical Center	**Nurse-managed community health center and a rural community-based training program** providing affordable and accessible health care
Open Arms Children's Health program	**Partnership between a comprehensive child-caring agency and a research university to offer on-site dental service and expanded health care access**
Mobile Health Services for Rural Kentucky (MHSRK) program	Increasing access to health care by providing **free screenings and affordable cardiac testing**, increasing knowledge of risk factors for heart disease, reducing unhealthy behaviors, and improving health outcomes among low-income, uninsured, and underserved rural residents
Meade Activity Center	**After-school program** to promote physical activities for children in a center
Community Health Worker (CHW) program	**Employing community health worker** to provide health care to reach a wider community
Get Healthy Access Program (GHAP)	**Program offers cardiovascular screenings**, including blood pressure, blood sugar, cholesterol, and BMI screenings and, for those eligible, enrollment in the GHAP program that offers health education and navigation to primary care providers
Hope Health Clinic	**Embedded patient navigation system** offers low-income, uninsured clients, families, and caregivers assistance to help overcome health care system barriers and facilitate timely access to services
St. Elizabeth Telepsychiatry program	**Remote access to psychiatric assessments via telemedicine**
Community Based Delivery Model: Virtual Care	**Primary and specialty care through tele-health** to poor and uninsured populations
Social Innovation for Missouri	**Community Health Improvement model emphasizes collaboration across multiple sectors and among diverse stakeholders** to achieve a strong community-wide initiative via policy, environmental, and community changes to address tobacco use and obesity
Soccer for Success	**After-school program** provides children in underserved communities with access to quality physical activity programs and nutrition education to support their physical and personal development, improve fitness, and reduce obesity
Carrera Adolescent Pregnancy Prevention Program (CAS-Carrera)	**School-based adaptation** of an evidence-based program designed to develop young people's capacity and desire to avoid parenthood and other risky behaviors during adolescence, and help them break the cycle of poverty and despair affecting their full development

Corporation for National and Community Service
250 E Street, SW
Washington, D.C. 20024

TEL: (202) 606-5000
TTY: (202) 606-3722

www.NationalService.gov