

Investing in Innovation (i3): Strong Start on Evaluating and Scaling Effective Programs, But Greater Focus Needed on Innovation

Early-stage Grant Selection Should Be Improved, Overall Program Better Integrated with Other Initiatives

By Patrick Lester ¹
January 19, 2017

Executive Summary

This report is an evaluation of the Investing in Innovation (i3) program, a tiered-evidence grantmaking initiative at the U.S. Department of Education.² The program's primary purpose is to support the development, testing, and scaling of field-initiated programs for high-need students in K-12 education.

Created in 2009, it has provided over \$1.4 billion in grants for education projects, including those focused on kindergarten readiness, student achievement, decreasing dropout rates, and turning around low-performing schools. In late 2015, the program was changed as part of the Every Student Succeeds Act. However, the renamed Education Innovation and Research (EIR) program has retained most of i3's original features.

This report reviews the program's early progress. Its findings are based on a review of publicly available final project evaluations, internal performance reports obtained through a Freedom of Information Act (FOIA) request, and interviews with current and former officials from the U.S. Department of Education, i3 project directors, and several national experts in education.

The report includes an assessment of the program's overall results, its contributions to the knowledge base, and lessons learned from launching, implementing, evaluating, and scaling i3-funded projects. The remainder of this executive summary provides highlights from the full report.

Early Results

- **Evaluation Results:** As of January 1, 2017, final evaluations have been released for 44 i3 projects. Of these, 13 have positive impact findings (30 percent) and another seven (16 percent) produced mixed results, with positive effects reported on at least one measure.

As expected, a higher percentage of the program's scale-up and validation grants, which required more evidence, have produced positive impacts (50 percent). A smaller share of development grants, which required less evidence, did so (20 percent). Although comparisons should be made cautiously, these rates of success appear to exceed those in other areas of education research.

- **Affected Issues in Education:** The top 13 evaluations, all of which are based on randomized controlled trials (RCTs) or quasi-experimental designs (QEDs), have demonstrated positive effects for programs in reading and literacy, kindergarten readiness, STEM (science, technology, engineering, and math), the arts, charter schools, distance learning in rural communities, college preparation, and teacher professional development.
- **Evaluation Pipeline:** If the current rate of positive impact findings is sustained (30 percent), a total of 52 final evaluations with positive results

¹ For more information, contact Patrick Lester, Director, Social Innovation Research Center, at (443) 822-4791 or patrick@socialinnovationcenter.org.

² Andrew Feldman and Ron Haskins, "Tiered-Evidence Grantmaking," September 9, 2016. Available at: <http://www.evidencecollaborative.org/toolkits/tiered-evidence-grantmaking>

will be generated from the 172 grants that have been made under the program (2010-2016), or four times the number of final evaluations with positive impact results (13) that have been released to date.

- **Scaling Evidence-based Initiatives:** Results have been released for four scale-up grants, the largest of the i3 grants which are intended to expand programs backed by the highest levels of evidence. All four scale-up grantees – KIPP, Teach for America, Success for All, and a Reading Recovery program launched by Ohio State University – expanded their evidence-based programs, although some missed their self-identified growth targets.

Two expanded with positive impact findings in their evaluations, while the other two did so with mixed findings.

These results appear to be aligned with earlier research that suggests that strong intermediaries may be needed to successfully scale evidence-based programs in low-performing schools. As a group, they performed better than local school districts that also received i3 grants, but acted largely on their own.

Recommendations

While the i3 program (now EIR) appears to be achieving many of its intended objectives, it could be improved in the following ways:

- **EIR Should Rework Its Early-phase Grants to Better Support Genuine Innovation:** While i3-funded projects have produced positive effects at higher rates than has been typical in education research, its support for new and innovative programs appears to be one of its weakest features. Such projects were supported through the program's lowest-tier grants. While some of these grants have produced positive results, they appear to have generated few, if any, groundbreaking innovations.

The new EIR program has taken steps to address this issue by being more supportive of flexibility and continuous improvement in the early-phase grants, but more is needed. The selection process for these grants should be reworked, with greater reliance on national experts who are aware of gaps in existing research and can more readily identify true innovations. Early-phase grantees should also be

offered more tailored technical assistance that better connects them to experts in their respective fields of interest.

- **EIR Should Support Faster Research:** Final evaluation results for most of the first-year grants, which were awarded in 2010, did not become available until 2016. While some research takes more time, six years is too long to wait for results in most cases.

Much of this delay has been due to the program's simultaneous scaling expectations, which create delays as new staff are hired and new initiatives are launched in new schools.

The pace of research could be hastened for early-phase and mid-phase grants by providing more grants to programs that already have operations underway in multiple schools and do not require further expansion. The program should also offer lower-cost, short-duration grants like those that have been funded by the Institute of Education Sciences.

- **EIR Should Connect to and Leverage Other Publicly-funded Education Programs:** As noted earlier, the first cohort of scale-up grantees expanded their programs with either positive or mixed effects. As also noted, one major lesson of these efforts seems to be that successfully scaling evidence-based programs may require the involvement of high-capacity intermediaries like those that have been funded by i3.

To date, demand for evidence-based programs and models has been weak, but the Every Student Succeeds Act has laid the groundwork for increased use of evidence through several of its provisions, including reworked state accountability measures and new evidence definitions that apply to formula-funded and competitive grant programs. The Department of Education is providing guidance to states and local school districts on how to implement the evidence provisions of the new law.

Given the increased importance of these efforts, the limited size of i3's (now EIR's) budget, and the apparent importance of high-capacity intermediaries, the Department may wish to consider ways to better integrate EIR with these other efforts by providing incentives to applicants that can leverage other federal, state, and local program funds.

Contents

Executive Summary	1
Introduction	4
Chapter 1 – Early Results	
Summary Results	6
Progress in Priority Areas	9
Reasons for Success or Failure	13
Chapter 2 – Launch	
Obtaining the Grant	17
Program Launch	19
School Partnerships	20
Capacity Building	22
Chapter 3 – Innovate	
Innovation	25
Evaluation	29
Data	34
Chapter 4 – Scale	
Sustainability	42
Dissemination	44
Scale	45
Epilogue	50
Recommendations	55
Appendices	
Project Summaries	59
Links to Final Evaluations	64

Introduction

This report is an evaluation of the Investing in Innovation (i3) program, an evidence-based education initiative housed within the Office of Innovation and Improvement (OII) at the U.S. Department of Education.³ Its primary purpose is to support the development, testing, and scaling of effective, field-initiated programs that support growth and academic achievement for high-need students in K-12 education.

Created in 2009 during the first months of the Obama administration,⁴ it has made over \$1.4 billion in grants⁵ and created a portfolio of projects covering a range of issues such as improving literacy, closing achievement gaps, decreasing dropout rates, and increasing college enrollment and completion.

Evidence in Education

The i3 program is part of a larger national effort to build and increase the use of evidence-based programs and practices in education. This broad effort has included substantial investments in both education research and the dissemination of research findings to national, state, and local educators and policymakers.

At the federal level, most of this work is housed at the Institute of Education Sciences (IES) at the U.S. Department of Education, which has an annual budget of over \$500 million.⁶ Its two research arms are the National Center for Education Research (NCER) and National Center for Special Education Research (NCSEER).

Dissemination efforts are led by another IES division, the National Center for Education Evaluation and Regional Assistance (NCEE). This center oversees the What Works Clearinghouse (WWC), which reviews and rates existing research and makes its reviews and broader summaries available to the public through its web site. The center also houses the Education Resources Information Center (ERIC), an online library of research and information, and the Regional Education Laboratories (RELs), which work with school districts, states, and others to support the practical application of evidence-based practices.

In addition to these federal efforts, states and local school districts also play a major role, including conducting research and overseeing the implementation of evidence-based programs in their respective jurisdictions.

The i3 Program

The i3 program complements these broader efforts, but with some important differences. While much education research is led by academics,⁷ most i3 grants are made to local practitioners, either local school districts or nonprofits that are working in partnership with them.⁸

Through its grants, the i3 program has attempted to create a “pipeline” of projects, with each operating at one of three different tiers of development – early-stage innovations, mid-level programs with some evidence, and initiatives with substantial evidence that should be expanded nationally.⁹ Each of these tiers is addressed by one of three different types of

³ Additional details about i3 can be found on the program’s web site at <https://www2.ed.gov/programs/innovation/index.html> and at the i3 Learning Community at: <https://i3community.ed.gov/>

⁴ Authorizing provisions were included in section 14007 of the American Recovery and Reinvestment Act of 2009 (PL 111-5). See: <https://www2.ed.gov/policy/gen/leg/recovery/statutory/stabilization-fund.pdf>

⁵ Communication with OII staff, January 10, 2017.

⁶ More information about IES is available on its web site at: <https://ies.ed.gov/>

⁷ Academically-led research has been criticized for sometimes failing to meet the needs of practitioners. For a discussion of these issues, see: Thomas Kane, “Connecting to Practice: How We Can Put Education Research to Work,” *Education Next*,

2016. Available at: <http://educationnext.org/connecting-to-practice-put-education-research-to-work/>; Grover J. “Russ” Whitehurst, “Evidence In Education: A Look to the Future,” *Education Next*, April 16, 2014. Available at: <http://educationnext.org/evidence-education-look-future/>; Ruth Curran Neild, “Federally-supported Education Research Doesn’t Need a Do-over,” Brookings Institution, April 7, 2016. Available at: <https://www.brookings.edu/research/federally-supported-education-research-doesnt-need-a-do-over/>; and GAO, “Education Research: Further Improvements Needed to Ensure Relevance and Assess Dissemination Efforts,” December 5, 2013. Available at: <http://www.gao.gov/products/GAO-14-8>

⁸ Some i3 nonprofit grantees are universities, but this is a small fraction of total grants.

⁹ Alyson Klein and Sarah Sparks, “Investing in Innovation: An

program grant – development, validation, and scale-up.¹⁰

Under i3, the smaller development grants are intended for new innovations and have the lowest incoming evidence requirements. The largest scale-up grants have the highest evidence requirements and are expected to expand their programs to many new schools and communities. The mid-tier validation grants fall in between. All three types of grant are also expected to conduct an independent evaluation to determine their effectiveness.

In 2015, Congress updated the program as part of the Every Student Succeeds Act.¹¹ The new law added state education agencies, tribes, and other organizations to the list of eligible applicants, but otherwise most of the essential features of the newly renamed Education Innovation and Research (EIR) program remain unchanged.¹² It continues to support a broad portfolio of K-12 projects operating at three different points in the development process: (1) early-stage innovations; (2) mid-phase projects with more rigorous evaluations; and (3) the expansion or scaling of effective programs.

This report reviews the progress of this program toward accomplishing these objectives.

Methodology

This report is based on several sources of information. They include a review of publicly available final evaluations and internal performance reports obtained through a Freedom of Information Act (FOIA) request.¹³

The report is also based on interviews with current and former officials from the U.S. Department of Education,¹⁴ i3 technical assistance providers, 65 project directors representing 69 out of the 117 grants made by the i3 program from 2010-2013,¹⁵ and several national experts in education and evidence-based policy.¹⁶

Except where views are attributed by name, the opinions expressed in this report are not necessarily endorsed or shared by these individuals or organizations.

Plan of the Report

This report provides an overview of the i3 program's progress and factors that have contributed to its performance. It is organized as follows:

- **Chapter One** reports on i3's early results. It also discusses major factors that contributed to the success or failure of individual i3 projects.
- **Chapter Two** reviews early project activities, including obtaining the grant, project launch, school partnerships, and capacity building.
- **Chapter Three** reviews experiences with innovation, evaluation, and data.
- **Chapter Four** provides insights on sustainability, dissemination of project results, and scale.
- The **Epilogue** provides expert reactions to the report's findings and opinions on the program's future and role in education more broadly.

The report concludes with recommendations.

Introduction to i3, *Education Week*, March 22, 2016. Available at: <http://www.edweek.org/ew/articles/2016/03/23/investing-in-innovation-an-introduction-to-i3.html>

¹⁰ These tiers have been renamed under the EIR program, which replaced i3. They are now called early-phase, mid-phase, and expansion grants.

¹¹ Social Innovation Research Center, "K-12 Education Bill Advances Evidence-based Policy, Replaces i3," December 7, 2015. Available at: <http://www.socialinnovationcenter.org/?p=1806>

¹² Social Innovation Research Center, "ED Announces First Round of Grants Under i3's Replacement," December 15, 2016. Available at: <http://www.socialinnovationcenter.org/?p=2430>

¹³ FOIA Request No. 16-00347-F was submitted on November 14, 2015 and completed in full on March 21, 2016. The request obtained the latest performance report, final evaluation, or both

for all 117 i3 grantees receiving awards during the 2010-2013 program years.

¹⁴ Interviews were conducted with staff of the Office of Innovation and Improvement, i3 program staff, and officials at the Institute of Education Sciences. These interviews were conducted in 2016 and early 2017. Most were political appointees. At the time of the interviews most were still serving in an official capacity, but most have now left the Department.

¹⁵ Project director interviews were conducted from June 10 to August 1, 2016. Some interviews included additional project representatives, such as the program evaluator. Four interviews were conducted with organizations with two i3 grants from the 2010-2013 years.

¹⁶ All individuals who were interviewed were given an opportunity to review the draft report, make comments, and offer corrections.

Chapter One: Early Results

How well is the i3 program working? What is it accomplishing?

This chapter summarizes: (1) results for project evaluations that have been released publicly so far; (2) how these results fit into the broader body of knowledge in their respect focus areas; and (3) factors that contributed to the success or failure of individual i3 projects.

Summary Results

Although the results have varied by grant tier, evaluation methodology, and project focus (see **Tables 1 and 2**), just under a third of the 44 projects with final evaluations have generated positive impact results so far.

- **Project Success Rates:** Of the 44 projects with publicly-released final evaluations as of January 1, 2017, 13 have generated positive impact results (30 percent)¹⁷ and another seven (16 percent) produced mixed results that included positive results on at least one impact measure. Another 18 projects (41 percent) generated no impact and the remaining six conducted evaluations that generated only preliminary evidence (14 percent).¹⁸

These results should be interpreted with some caution, however, because they are largely based on findings as reported in the independent evaluations.¹⁹ Subsequent independent reviews, like those commonly conducted by evidence clearinghouses, sometimes question a study's underlying methodology. However, according to the Department of Education, all of the validation

and scale-up grants and most of the development grants are on track to meet What Works Clearinghouse standards.²⁰

- **Results by Grant Tier:** Projects funded by larger scale-up and validation grants were more likely to achieve positive impact results (50 percent for each category) than projects in the lower tier development grants (20 percent).²¹

The lower percentage of positive impact results among development grantees was expected because of their lower evidence thresholds to receive a grant. However, better results among scale-up and validation grants came despite facing more significant scaling requirements, a factor that can often undermine results (and is described further in Chapter Four).

- **Results by Evaluation Methodology:** Projects that conducted randomized-controlled trial (RCT) studies were about as likely to have positive results (35 percent) as those using quasi-experimental designs (33 percent). Among grantees that used RCTs, the equivalent rate for developmental grants (27 percent) was lower than that for validation (37 percent) and scale-up grants (50 percent).

These rates of success appear to meet or exceed those experienced in other education research efforts. A 2013 Coalition for Evidence-Based Policy review of well-conducted randomized controlled trials commissioned by IES found that only 12 percent found positive effects.²²

¹⁷ Summaries of the 13 projects that produced positive impact can be found in Appendix A. Links are also included to What Works Clearinghouse study reviews where they are available.

¹⁸ Links to the all final evaluations, including those that generated mixed or no impact, can be found in Appendix B.

¹⁹ Of the 44 final evaluations rated in this report, WWC study reviews were publicly available for 13. SIRC ratings and WWC reviews are aligned on those 13 evaluations. The other 31 were rated by SIRC based on the findings as reported in the evaluations. Links to WWC reviews, where available, are included in Appendix A.

²⁰ U.S. Department of Education, "Innovation and Improvement: Fiscal Year 2016 Budget Request," pp. G-24-26. Available at: <https://www2.ed.gov/about/overview/budget/budget16/justifications/g-ii.pdf>

²¹ A senior Department of Education official overseeing the program has said that the success rate for development grants is high, exceeding the average for venture capital projects. See Sarah Sparks, "Lessons From i3: California, Georgia Schools Learn From 'Failed' Interventions," *Education Week*, March 24, 2016. Available at: http://blogs.edweek.org/edweek/inside-school-research/2016/03/stories_from_i3_california_sch.html

²² Coalition for Evidence-Based Policy, "Randomized Controlled Trials Commissioned by the Institute of Education Sciences Since 2002: How Many Found Positive Versus Weak or No Effects," July 2013. Available at: <http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf>

Table 1: The Innovation Pipeline – Current Status of i3 Final Evaluations

The following table shows the current status of final evaluations for i3's 2010-2013 grantees as of January 1, 2017. Final evaluations are not yet available for grants awarded in later years.

Summaries for projects with positive impact can be found in Appendix A. Links to all of the publicly available final evaluations can be found in Appendix B.

<u>Grantees</u>	<u>2010</u>	<u>2011</u>	<u>2012</u>	<u>2013</u>	<u>Total</u>
Grants Awarded	49	23	20	25	117
o Final evaluations completed	39	4	1	0	44
o Final evaluation not yet available	10	19	19	25	73
Evaluation Results	39	4	1	0	44
o Positive impact	12	1	0	0	13
o Mixed impact	6	1	0	0	7
o No impact	17	1	0	0	18
o Preliminary evidence	4	1	1	0	6

Definitions

- o *Positive impact* = Positive results on half or more of important reported impact measures (statistically significant, substantive).
- o *Mixed Impact* = Positive results reported on at least one, but fewer than half of important impact measures.
- o *No impact* = No reported positive impact results.
- o *Preliminary evidence* = Evaluation with no substantially similar comparison group.

Source: SIRC ratings for final evaluation results are based on an analysis of publicly available final evaluations and What Works Clearinghouse study reviews. SIRC ratings are aligned with WWC reviews for the 13 evaluations where they are publicly available (links to WWC reviews can be found in Appendix A).

In other cases, ratings are based on SIRC interpretation of findings as stated in the evaluations and do not reflect a detailed review of the underlying evaluation methodology.

Table 2: Evaluation Results by Type of Grant and Evaluation

The following table shows results for the 44 final evaluations broken out by type of grantee and evaluation.

Grantees	Number	Positive Impact	Mixed Impact	No Impact	Preliminary Evidence	Percent Positive
Total Evaluations	44	13	7	18	6	30%
Results by grant type						
o Scale-up	4	2	2	0	0	50%
o Validation	10	5	3	2	0	50%
o Development	30	6	2	16	6	20%
Results by absolute priority						
o School turnarounds	9	4	1	2	2	44%
o Standards and assessment	15	6	3	6	0	40%
o Teacher / principal effectiveness	12	3	3	3	3	25%
o Data driven instruction	6	0	0	6	0	0%
o Other	2	0	0	1	1	0%
Results by evaluation type						
o Randomized controlled trials	23	8	5	10	0	35%
o Quasi-experimental designs	15	5	2	8	0	33%
o Preliminary evidence	6	0	0	0	6	0%
Development grants						
o Randomized controlled trials	11	3	0	8	0	27%
o Quasi-experimental designs	13	3	2	8	0	23%
o Preliminary evidence	6	0	0	0	6	0%
Validation grants						
o Randomized controlled trials	8	3	3	2	0	37%
o Quasi-experimental designs	2	2	0	0	0	100%
o Preliminary evidence	0	0	0	0	0	-
Scale-up grants						
o Randomized controlled trials	4	2	2	0	0	50%
o Quasi-experimental designs	0	0	0	0	0	-
o Preliminary evidence	0	0	0	0	0	-
Local school districts						
o Randomized controlled trials	6	0	1	5	0	0%
o Quasi-experimental designs	6	3	0	3	0	50%
o Preliminary evidence	4	0	0	0	4	0%

Source: SIRC ratings for final evaluation results. For rating definitions, see Table 1. Summaries of projects with positive impact, including links to What Works Clearinghouse reviews where available, can be found in Appendix A. Links to the final evaluations for all of the projects are in Appendix B.

As noted earlier, however, comparisons should be made cautiously because third-party reviews of the underlying methodology are publicly available for only some of the i3 evaluations.²³

Six of the other project evaluations only contained outcomes information. Four of these tracked outcomes over time. Another provided limited comparisons across schools, but there was poor baseline equivalence between the two groups. The sixth was a qualitative study. Of the first five, three showed no apparent program effects and two showed mixed effects. Four of the six were evaluations of projects run by local school districts.

- **Results for Local School Districts:** Of the 16 grants made to local school districts, three (19 percent) have generated positive impact results. All but one of these 16 were development grants and the success rate for this group is comparable to that for development grantees overall (20 percent).

As a group, the local school districts appeared to have some advantages and disadvantages compared to the other grantees. Advantages were tied to their location in the schools. They were often closer to the work, had an easier time with buy-in, easier access to data, and budgets that could sustain a program if it was working (although often these budgets were substantially constrained). Disadvantages included lower capacity in some critical areas of expertise (especially in evaluations), less specialization or experience with the chosen intervention, and a lack of direct access to national experts.

In their evaluations, the success rates for local school districts varied according to the methodology used. Among the six that conducted randomized controlled trial (RCT) studies, none produced positive impact results (one produced mixed results). Four others produced evaluations with only preliminary evidence (i.e., evaluations with no substantially similar comparison group). As noted earlier, most of the grantees that produced only preliminary evidence (4 of 6) were local school districts.

Schools or school districts that used quasi-experimental designs (QEDs), which compared

results for program participants to others within the same school or at matched schools, were more likely to generate positive impact results (3 of 6).

- **Results for Universities:** Four of the grants with final evaluations were made to universities. All four were either validation (3) or scale-up grants (1). Of these, three achieved positive impacts and one achieved mixed impact results.²⁴ None of them produced no impact.
- **Results for Scale-up Grants:** Scaling effective programs was one of i3's central goals. All four of the 2010 scale-up grantees expanded their programs. Two did so while generating positive impacts. The other two did so with mixed results. These results, and associated lessons learned, are discussed in Chapter Four.
- **Overall Progress:** A total of 172 grants have been awarded under the program from 2010-2016.²⁵ Of these, final evaluations have only been released from the first three years (2010-2012) and a large majority of those are from the program's first year (2010). Final evaluations are now available for about 80 percent of the grants from that year (39 of 49).
- **Evaluation Pipeline / Projection:** If the current success rate is sustained (30 percent), a total of 52 final evaluations with positive impact results will be generated from the 172 grants that have been made under the program (2010-2016), or four times the number of final evaluations with positive impact results (13) that have been released to date.

Progress in Priority Areas

The successful i3 projects have produced positive findings across a variety of important education issues, including reading and improved literacy, STEM, and kindergarten readiness, among others. This section reviews those results in more detail.

All but two of the 44 final evaluations released so far fall into one of the following four categories: (1) school turnarounds, (2) standards and assessments, (3) teachers and principals, and (4) use of data.²⁶

²³ As noted previously, links are provided to WWC reviews where they are available in Appendix A.

²⁴ The three with positive impact results were: the Ohio State University scale-up grant and the Utah State University and

University of Missouri validation grants.

²⁵ A list of current awards can be found on the i3 web site at <https://www2.ed.gov/programs/innovation/awards.html>

²⁶ These four categories were absolute priorities for i3 in its first

Results for each of these categories (called absolute priorities) are summarized in **Table 2**.

Among these four absolute priorities, the highest success rates were experienced in the school turnaround and standards and assessment groups. Fewer successes were experienced in the teacher and principals group, which is focused primarily on professional development, or among the data-focused grantees. The highest success rate of all was experienced by grantees that were focused on reading and literacy, although this group lacked a formal designation and they were spread out among the other categories.

These evaluation results are discussed in further detail below. More detailed descriptions of the mentioned projects can be found in Appendix A.

School Turnarounds

Turning around low-performing schools has been a focus for federal policymakers dating at least back to the enactment of No Child Left Behind. Efforts to turn around these schools received a boost in 2009 when Congress authorized \$3 billion for School Improvement Grants (SIG) as part of the American Recovery and Reinvestment Act, the same law that created i3. The SIG program continued to receive about \$500 million per year in funding after that.

The SIG program promoted a number of specific reform strategies, including school closures, conversion into charter schools, replacing teachers and principals, and adopting other reforms such as

merit pay for teachers.²⁷ Many of these were controversial²⁸ and the program's overall effectiveness was modest.²⁹

SIG was eliminated by the Every Student Succeeds Act (ESSA) and replaced with a state set aside for school improvement under Title I.³⁰ Under the new law, the Department of Education may not mandate specific strategies. States and local school districts have greater flexibility, so long as their chosen strategies are backed by evidence.³¹

The number of evidence-backed school turnaround models is small, however.³² One of them is Success for All, a whole school reform strategy that includes job-embedded professional development and coaching, collaborative performance monitoring, curriculum resources, and strategies for addressing school-wide issues such as low attendance. Previous evaluations have showed that students in SFA schools achieve better academic outcomes, including in reading.³³

Under its i3 grant, Success for All (SFA) scaled up its model in 600 elementary schools. In this case, however, the evaluation results fell somewhat short of the earlier research, with positive effects reported on student phonics and pre-literacy skills, but no effects on reading comprehension, special education designations, or rates at which students were held back to repeat a grade.³⁴

SFA's scale-up may have faced challenges due to timing. It was rolled out in the immediate aftermath of the 2008-2009 recession, a period when many schools were facing budget shortfalls, and its i3 study

year (2010), which explains their predominance among the evaluations that have been released so far. The two evaluations that did not fall into one of these four categories were from grants made in later years (2011 and 2012). They were focused STEM and parent and family engagement respectively.

²⁷ Alyson Klein, "School Improvement Grant Efforts Face Hurdles," *Education Week*, April 25, 2011. Available at: <http://www.edweek.org/ew/articles/2011/04/27/29sig.h30.html>; Government Accountability Office, "Education Should Take Additional Steps to Enhance Accountability for Schools and Contractors," April 12, 2012. Available at: <http://gao.gov/products/GAO-12-373>

²⁸ Jaclyn Zubrzycki, "School Shutdowns Trigger Growing Backlash," *Education Week*, October 16, 2012. Available at: http://www.edweek.org/ew/articles/2012/10/17/08closings_ep.h32.html

²⁹ Alyson Klein, "Data Paints Mixed Picture of Federal Turnaround Program," *Education Week*, December 1, 2015. Available at: <http://www.edweek.org/ew/articles/2015/12/02/new-data-paints-mixed-picture-of-federal.html>

³⁰ Alyson Klein, "ESSA Clears Out Underbrush on School Improvement Path," *Education Week*, September 27, 2016. Available at:

<http://www.edweek.org/ew/articles/2016/09/28/essa-clears-out-underbrush-on-school-improvement.html>

³¹ Alyson Klein, "How Will ESSA Be Different When it Comes to School Turnarounds Than SIG?" *Education Week*, October 25, 2016. Available at: <http://blogs.edweek.org/edweek/campaign-k-12/2016/10/essa-different-sig-school-turnarounds.html>; Daarel Burnette II, "States, Districts to Call Shots on Turnarounds Under ESSA," *Education Week*, January 5, 2016. Available at:

<http://www.edweek.org/ew/articles/2016/01/06/states-districts-to-call-shots-on-turnarounds.html>

³² U.S. Department of Education, "Approved Evidence-Based, Whole-School Reform Models." Available at <https://www2.ed.gov/programs/sif/sigevidencebased/index.html>

³³ Geoffrey D. Borman, Robert E. Slavin, Alan C. K. Cheung, Anne M. Chamberlain, Nancy A. Mad-den, and Bette Chambers, "Final Reading Outcomes of the National Randomized Field Trial of Success for All," *American Educational Research Journal*, 44, no. 3 (2007): 701-731. Available at: <http://files.eric.ed.gov/fulltext/ED485351.pdf>

³⁴ MDRC, "Scaling Up the Success for All Model of School Reform," September 2015. Available at: <http://www.mdrc.org/publication/scaling-success-all-model-school-reform>

found that resource constraints had prevented some schools from faithfully implementing some key program components, including hiring a full-time facilitator or using SFA's computerized tutoring program.³⁵

Another i3-supported whole school turnaround effort is Diplomas Now, launched by Johns Hopkins University with a 2010 validation grant. The program is working in 32 middle and high schools to increase high school graduation and college readiness. While the project is not yet complete (and its results are therefore not included in Tables 1 or 2), its early results are still noteworthy. According to an interim RCT-based evaluation, after two years it reduced the percentage of students exhibiting one or more early warning signs that a student will drop out, including poor behavior, low attendance, or poor academic performance.³⁶

Smaller i3 development grants have also been used for whole school transformation efforts, but of the two with final evaluations that have been released so far, neither has had any effect. One that was conducted by a local school district was poorly implemented. The other was well-implemented, but failed to affect student test scores. While it is still early, these experiences, coupled with the widespread challenges faced under the School Improvement Grant program, suggest that successful whole school reform might be too much to ask of smaller i3 grantees (such as individual local schools or school districts) that are working without substantial and experienced outside support.

Compared to the whole school turnaround efforts, the other more targeted efforts were more successful, with all four final evaluations with positive impacts in the broader category coming from this subgroup. One was the Building Assets-Reducing Risks (BARR) turnaround project, which received a development grant through the Search Institute. This project provided targeted support for 9th graders by organizing students into cohorts of 30 and providing a variety of professional development and family engagement supports. It successfully boosted achievement in a school in suburb outside Los

Angeles, according to its RCT-based study.

Another successful project was in New Mexico, where a validation grant was used to support a kindergarten readiness and academic achievement initiative for K-3 students, called StartSmart K-3 Plus. The program's RCT-based study showed that it improved vocabulary skills for pre-K students and also increased reading, math, and writing test scores for both these students and for older students up through the start of third grade.

Two other successful projects in this category were literacy-related, the Milwaukee Community Literacy Project and the Reading Recovery scale-up grant sponsored by Ohio State University. They are discussed again later in this section.

Standards and Assessments

As a group, grantees in the standards and assessment absolute priority performed well, Out of 15 grantees with final evaluations, six produced positive impacts and three more generated mixed impacts.

Of the six with positive impacts, three were focused on STEM (science, technology, engineering, and math). One was a grant to the University of Missouri to test an inquiry-oriented professional development program focused on improving student math and English skills. The second was a STEM professional development project overseen by ASSET, Inc. The third was a STEM-focused career and college readiness project in the Bellevue School District in Washington state. (For details, see Appendix A.)

Student scores on the National Assessment of Education Progress in math and science have leveled off in recent years.³⁷ They also continue to show significant racial and gender gaps. Some argue that these differences have also increased high school dropout rates beyond what they would have been.³⁸

While there are many potential contributors to the success of STEM initiatives in general, two issues

³⁵ Ibid., pp 114-116. The MDRC evaluation of Success for All found that implementing the program cost schools an extra \$227 per student per year, including additional staff-related costs, or about \$168,348 per school. These figures are separate from national costs incurred by SFA itself, including those supported by the i3 grant.

³⁶ MDRC, "Addressing Early Warning Indicators: Interim Impact Findings from the Investing in Innovation (i3) Evaluation of Diplomas Now," June 2016. Available at: <http://www.mdrc.org/project/diplomas-now>

³⁷ National Center for Education Statistics, "The Nation's Report Card." Available at: <https://www.nationsreportcard.gov/>; Marva Hinton, "Science Scores Rise for 4th and 8th Grades," *Education Week*, October 27, 2016. Available at: <http://www.edweek.org/ew/articles/2016/10/27/science-scores-rise-for-4th-and-8th.html>

³⁸ Andrew Hacker, "Is Algebra Necessary?" *The New York Times*, July 28, 2012. Available at: <http://www.nytimes.com/2012/07/29/opinion/sunday/is-algebra-necessary.html>

have stood out: severe shortages in qualified math and sciences teachers³⁹ and the need to shift to a more active, hands-on approaches to teaching.⁴⁰ At least one of these two issues was addressed by each of the successful i3 STEM grants, either through professional development, new teaching techniques, or both.

The i3 program has since increased its focus on STEM, making it a priority for grants awarded from 2011-2015. There are now more than a dozen additional STEM-related projects in the i3 pipeline.

The other three successful non-STEM related grantees in the standards and assessments category were: The Studio in a School, a project that developed open access education resources and assessments in the arts; the Fresno County Office of Education, which developed a reading and writing course for high school seniors that produced improved test scores; and the Niswonger Foundation, which developed a successful distance learning program that improved ACT and advanced placement test scores in rural Tennessee.

Teachers and Principals

Twelve of the i3 projects with final evaluations were focused on supporting effective teachers and principals. All twelve either focused on or included professional development, but only three generated positive impact findings and one of these, the KIPP charter school scale-up, was only tangentially about professional development.

The other two were both reading related: the Children's Literacy Initiative, which provided literacy instruction for K-3 teachers, and the Iredell-Statesville Schools in North Carolina, which tested a professional development initiative that raised

reading test scores for students in grades 3-8.

Outside of KIPP and the two reading-related projects (which as a group did well and are discussed below), none of the other professional development-focused projects in this category generated positive impact results.

However, two of these grantees – Teach for America and the IDEA Public Schools in the Rio Grande Valley in Texas – were arguably close. Both ran programs that recruited college graduates and professionals with strong academic backgrounds and leadership experience to work in low-performing public schools. Neither program achieved positive impact because their evaluations showed little difference in student test scores when compared to incumbent teachers. However, in both cases the teachers in the comparison group on average had substantially more experience, which suggests that the comparisons may have been unfair.

The poor overall performance in this category may not be surprising. Previous studies have suggested that most teacher professional development is ineffective⁴¹ and does not meet quality standards established under ESSA.⁴² Several national organizations are working to address this problem, however, including the Council for the Accreditation of Educator Preparation, which has been moving toward evidence-informed accreditation.⁴³ Under ESSA, grants under the Supporting Effective Educator Development (SEED) grant program, which is run by the same office that oversees i3, must also be evidence-based.⁴⁴ Better connections between grantees and these efforts may generate better results for future EIR professional development grants.

³⁹ Michael Marder, "Is STEM Education in Permanent Crisis?" *Education Week*, October 25, 2016. Available at: <http://www.edweek.org/ew/articles/2016/10/26/is-stem-education-in-permanent-crisis.html>; Kirsten Daehler, "The Key to Good Science Teaching," *Education Week*, October 25, 2016. Available at: <http://www.edweek.org/ew/articles/2016/10/26/the-key-to-good-science-teaching.html>

⁴⁰ Anne Jolly, "How to Design a Successful STEM Lesson," *Education Week*, September 28, 2016. Available at: <http://www.edweek.org/tm/articles/2016/09/23/how-to-design-a-successful-stem-lesson.html>; Kirsten Daehler, "The Key to Good Science Teaching," *Education Week*, September 28, 2016. Available at: <http://www.edweek.org/ew/articles/2016/10/26/the-key-to-good-science-teaching.html>; and Mike Schmoker, "Math and K-12 Schools: Addressing the Historic Mismatch," *Education Week*, September 28, 2016. Available at:

<http://www.edweek.org/ew/articles/2016/08/12/math-and-k-12-schools-addressing-the-historic.html>

⁴¹ Madeline Will, "Study: Most Professional Training for Teachers Doesn't Qualify as 'High Quality'," *Education Week*, November 23, 2016. Available at: http://blogs.edweek.org/teachers/teaching_now/2016/11/essa_pd_report.html

⁴² Stephen Sawchuk, "Study Casts Doubt on Impact of Teacher Professional Development," *Education Week*, August 18, 2015. Available at: <http://www.edweek.org/ew/articles/2015/08/19/study-casts-doubt-on-impact-of-teacher.html>

⁴³ Emerson J. Elliott, "Promoting Evidence-based Teacher Preparation," February 4, 2016. Available at: <http://wtgrantfoundation.org/evidence-crossroads-pt-9-promoting-evidence-based-teacher-preparation>

⁴⁴ Information on the SEED grant program is available at: <https://www2.ed.gov/programs/edseed/index.html>

Data Driven Instruction

Data-driven instruction holds the potential for improving student outcomes by allowing teachers to: (1) identify students with gaps in knowledge or achievement; and (2) provide targeted and timely instruction. So far, however, there has been limited causal evidence of this strategy's effectiveness.⁴⁵

Six of the i3 grants with final evaluations were focused on the use of data, but all six produced no impact, making it the lowest-performing of the four main absolute priorities. The reasons for this poor performance varied. One project was undermined by problems with the evaluation design (the evaluator could not create or find an appropriate control group). Another suffered problems with its IT vendors, which prevented it from implementing the program with fidelity during the study period. More broadly, data use in each of these projects was usually just one of many program components and it may have had little effect because it was embedded within larger programs that were ineffective.

While not definitive, one of the six grantees produced results that lent support to this view. Over the course of its grant, the Achievement Network supported data-driven instruction in 481 schools. Like the other five data grantees, its evaluation found no impact on student achievement in math or reading.

However, there was a bright spot in the evaluation results. Its study found that the program generated significant positive effects in schools that were assigned higher readiness ratings prior to the intervention. In these schools, educators both analyzed data more frequently and then used that analysis to shape their instruction.

By contrast, the program generated null results in schools assigned lower readiness ratings, where educators analyzed data more frequently but did not act on it to shape instruction. The results in these schools washed out the positive effects in the other schools, creating no effect overall.

The Achievement Network has since used these results to improve its program. It has worked to tailor

the initiative based on the readiness of its partner schools. It has also worked to make data use fit more seamlessly into the planning and instructional work of participating teachers.⁴⁶

Reading / Literacy

While not a formally designated absolute priority, at least seven i3 grantees developed projects related to reading or literacy. As a group, they had greater success than the others.

Of these seven projects, four achieved positive impact in their evaluations, two achieved mixed results, and one generated no impact. The four with positive results comprised almost a third of the thirteen i3 projects with positive results so far.

The four projects with positive impacts are mentioned above in their respective official categories. They include the Reading Recovery scale-up grant overseen by Ohio State University, a validation grant to the Children's Literacy Initiative for its Model Classrooms Project, and two development grants: one to the Boys and Girls Clubs of Milwaukee for their Milwaukee Community Literacy Project (SPARK) and the other to Iredell-Statesville Schools for their reading-focused teacher professional development initiative. The two with mixed results were the Success for All scale-up grant and a Reading Apprenticeship validation grant to WestEd.

Together, these results could inform a new program created under ESSA called Literacy Education for All, Results for the Nation (LEARN). The program authorizes grants for evidence-based literacy instruction in high-need schools.⁴⁷

Reasons for Success or Failure

Why did some projects succeed while others failed? ⁴⁸ Given their varied local contexts and multiple moving parts, this is a causal question that can be answered by this report only tentatively.

Nevertheless, based on a review of all 44 final evaluations, progress reports for the other 2010-2013

⁴⁵ Martin R. West, et al, "Achievement Network's Investing in Innovation Expansion: Impacts on Educator Practice and Student Achievement," March 2016, Harvard Center for Education Research, p. 1. Available at: <http://cepr.harvard.edu/achievement-network-evaluation>

⁴⁶ ANet, "i3 Study Takeaway 2: Data and Assessment are Critical Tools — But They Can Also Be Distracting," October 21, 2015. Available at: <http://www.achievementnetwork.org/anetblog/2015/10/21/lessons-from-our-i3-study>

⁴⁷ Liana Heitin, "ESSA Reins In, Reshapes Federal Role in Literacy," *Education Week*, January 5, 2016. Available at: <http://www.edweek.org/ew/articles/2016/01/06/essa-reins-in-reshapes-federal-role-in.html>; National Council of Teachers of English, "Literacy Advocates Heartened by LEARN Act's Inclusion in ESSA," December 18, 2015. Available at: <http://blogs.ncte.org/index.php/2015/12/literacy-advocates-heartened-by-learn-acts-inclusion-in-essa/>

⁴⁸ In this section, "success" and "failure" refers to positive evaluation findings except where otherwise specified.

grantees, and interviews with 65 of the project directors, the following factors seemed to make the greatest difference:⁴⁹

- **Difficulty of Program Objectives:** Projects that took on more difficult tasks seemed to be more likely to fail. One major factor was the project's chosen focus. For example, projects that focused on professional development appeared to face significant hurdles when judged according to their ability to affect student test scores, a finding that is consistent with the broader evidence in the field. Whole-school turnaround efforts also appeared to be more challenging, which is consistent with the broader experience in the federal School Improvement Grant program.⁵⁰

"My experience is that you will be more successful if you have a limited focus," said one grantee. "Whole school is hard, at least right off the bat."

Working with highly challenged, under-resourced schools was difficult even when grantees were not engaged in whole school turnaround efforts. This was particularly true in the early years of the Obama administration when many were subject to closures and school layoffs, either in response to turnaround efforts or because of severe budget shortfalls in the aftermath of the 2008-2009 recession.

- **Evidence-based or Well-Designed Interventions:** Initiatives with proven track records or strong evidence behind them seemed to be more likely to produce positive impact results. Evidence for this can be found in the higher positive impact rates among the scale-up and validation grants (see **Table 2**), which had to meet higher evidence standards to receive their grants.

Development grants, which faced lower evidence requirements, were not as likely to generate a positive impact. However, this is an expected trade-off for projects that are more innovative and developmental in nature.

- **Experience with the Chosen Intervention:** Organizations with substantial experience with their chosen intervention seemed to be more likely to succeed. This was especially true of grantees that had designed or substantially contributed to designing their interventions. Examples include all four scale-up grants, but also some of the validation and development grants, such as the BARR turnaround project.
- **Realistic Project Scale:** Scaling evidence-based programs presents substantial additional challenges (see the Scale section on in Chapter Four). These challenges were faced most clearly by the scale-up grants, but they were also faced to a lesser degree in many of the lower tier validation and development grants, some of which were engaged in substantial scaling activity.

In general, grantees that did less scaling and worked in fewer schools seemed to face fewer challenges. This may have been especially true for the truly new and innovative programs. One successful example was the Bellevue School District, which implemented its project in a single school and spent substantial time on qualitative and exploratory assessments before conducting its quantitative quasi-experimental study.⁵¹

Projects that focus on just one or relatively few schools may face other challenges in their evaluations, however, including the need for sufficient sample sizes for results to be statistically significant and generalizable. These tensions are discussed elsewhere in this report, including the sections on innovation and evaluation.

- **Sufficient Resources and Program Dosage:** Programs that attempt to spread limited resources too thinly (i.e., across too many schools, teachers, or students) may face substantial challenges. Evidence for this in i3 was not definitive, however, because the projects faced different cost structures depending on what they were doing. For some, like those that were developing new curricula, much of the cost was incurred up front and the subsequent per-pupil

⁴⁹ At least some information was obtained for all 117 grantees from the 2010-2013 grant years, either internal performance reports, final evaluations, or both. Interviews were conducted with project directors for more half of these, as described in the methodology section.

⁵⁰ Alyson Klein, "Data Paints Mixed Picture of Federal Turnaround Program," *Education Week*, December 1, 2015. Available at:

<http://www.edweek.org/ew/articles/2015/12/02/new-data-paints-mixed-picture-of-federal.html>

⁵¹ Randy Knuth, Ph.D., et al, "Re-imagining Career and College Readiness: STEM, Rigor, and Equity in a Comprehensive High School," March 1, 2016. Available at: <http://www.bsd405.org/wp-content/uploads/2016/03/Sammamish-i3-Grant-Findings-Report.pdf>

costs were comparably small. For others, contributions from the participating schools (financial or in-kind) could offset increased costs.

Nevertheless, stretching limited resources too far seemed to have had an effect in some cases. For at least two of the grantees, resource issues were clearly a major cause of program failure. In other cases, poor fidelity (see below) seemed to be a red flag. When programs were poorly implemented, it sometimes was because resources were insufficient to implement all of the program's requirements in the first place. In other cases, resources may have constrained a grantee's ability to take corrective action when fidelity problems were identified.

- **Grantee Capacity:** The internal capacity of i3 grantees seemed to be a factor in their success. What kinds of capacity mattered? Typically, they were the capacities mentioned elsewhere in this list, particularly intervention design, grantee experience, implementation fidelity, evaluation, and access to sufficient budgetary resources.

These capacity issues seemed to correlate with one another.⁵² In other words, high-capacity grantees seemed to experience fewer problems across the board, while low-capacity grantees seemed to face more problems across the board. Capacity was greatest among the large scale-up and validation grantees. It was more mixed among the development grantees, with school districts appearing to face the most severe capacity constraints.

In cases where the grantees were not school districts, their internal capacity appeared to be compensating for the low capacity of their partner schools (see below).⁵³ These high-capacity nonprofits seemed to be filling an intermediary or backbone function for the project.⁵⁴ Capacity seemed to play an especially large role for the scale-up grantees, which were working with many low-capacity schools at the same time (see the Scaling section of Chapter Four).

The i3 program contracted with two technical assistance providers to help address grantee capacity issues, Westat and Abt Associates. Their work is discussed in Chapter Two.

- **School Capacity and Buy-In:** The extent to which partner schools had sufficient capacity and were genuinely bought in to a project substantially influenced its chances for success. This is discussed in greater detail in the School Partnerships section of Chapter Two and the Scale section of Chapter Four.
- **Implementation Fidelity:** Failure to implement programs with fidelity was a problem for at least eight of the 44 projects, as determined by their reported performance on their chosen fidelity measures. Of these eight, four produced no impact in their impact studies, two were associated with studies that only tracked outcomes, and the last two reported positive impacts in their evaluations even though they missed their fidelity targets.

The quality of the chosen fidelity measures seems to have been an issue for several of the projects. For one of the two projects mentioned above (with positive impact results), the success thresholds seemed to be set too high (i.e., the project was demanding a level of fidelity that may have been unreasonable). For the other, the measures seemed to be poorly designed – the project achieved positive impact results overall even though its fidelity rating was very poor.

The quality of the fidelity indicators was also an issue in at least two other projects. Both technically met their fidelity targets, but generated poor impact results anyway. While these projects might have failed for other reasons, the fidelity measures seemed to be check-the-box affairs and did not adequately capture the true quality of implementation.

The quality of fidelity indicators seemed to be a widespread problem. For example, programs that provided professional development commonly

⁵² This judgment is based on performance reports obtained through the FOIA request, interviews with project directors, and information drawn from interim and formative evaluations, which typically reviewed implementation issues.

⁵³ This seemed to occur in the Social Innovation Fund too, where high-capacity SIF grantees seemed more able to address the challenges faced by their low-capacity sub-grantees. See Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, pp. 20-21, 30-38. Available at: [http://www.socialinnovationcenter.org/wp-](http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf)

[content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf](http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf)

⁵⁴ Backbone organizations play a central coordinating and support role in collective impact efforts, which bring together multiple organizations to solve a common problem. See Shiloh Turner, et al, "Understanding the Value of Backbone Organizations in Collective Impact: Part 1 in a 4 Part Series," *Stanford Social Innovation Review*, July 12, 2012. Available at: https://ssir.org/articles/entry/understanding_the_value_of_backbone_organizations_in_collective_impact_1

rated fidelity based on attendance or hours in class, rather than tests of acquired knowledge. This could suggest a greater level of teacher understanding than actually existed. These issues suggest that fidelity indicator design might be a topic that deserves additional attention from technical assistance providers.⁵⁵

- **Strength of Evaluation Designs:** Most of the i3 projects faced at least some challenges with their evaluations (see the Evaluation and Data sections of Chapter Three for a more detailed discussion). For at least ten of the 44 projects with final evaluations, however, poorly designed or implemented evaluations or data access issues were a major driver of poor results overall.

Seven of these ten were studies conducted for the 16 grantees that were local school districts. Of the broader group of ten, five were outcomes studies that included no equivalent comparison group. Three others used comparison groups in quasi-experimental designs (QEDs) that the project directors thought were inappropriate.

- **External Challenges:** Several of the grantees either failed or faced severe challenges for reasons that were outside their control. Examples included changes in the economy, related school budget problems, school closures and layoffs, changes in state tests, and the willingness of state and local education agencies to make data available to evaluators.

⁵⁵ Resources that may be helpful include: James Bell Associates, "Measuring Implementation Fidelity," October 2009. Available at: http://www.jbassoc.com/ReportsPublications/Evaluation%20Brief%20-%20Measuring%20Implementation%20Fidelity_Octob%E2%80%A6.pdf; Abby Hayes, et al, "Figuring out Fidelity: A Worked Example of the Methods Used to Identify, Critique and Revise the Essential Elements of a Contextualized Intervention in Health Policy Agencies," *Implementation Science*, February 24, 2016. Available at: <https://implementationscience.biomedcentral.com/articles/10.1186/s13012-016-0378-6>; Dennis Perez, et al, "A Modified

Theoretical Framework to Assess Implementation Fidelity of Adaptive Public Health Interventions," *Implementation Science*, July 8, 2016. Available at: <https://implementationscience.biomedcentral.com/articles/10.1186/s13012-016-0457-8>; and National Survey of Early Care and Education Project Team, "Measuring Predictors of Quality in Early Care and Education Settings in the National Survey of Early Care and Education," September 2015. Available at: <https://www.acf.hhs.gov/opre/resource/measuring-predictors-of-quality-in-early-care-and-education-settings-in-the-national-survey-of-early-care-and-education>

Chapter Two: Launch

What did it take to successfully start an i3 project? This chapter reviews several major components, including: (1) obtaining the grant; (2) program launch; (3) school partnerships, and (4) capacity building.

Obtaining the Grant

Applying for an i3 grant could be intimidating. Like most federal grants, the i3 application is loaded with bureaucratic requirements and jargon. Eligibility requirements, absolute and invitational priorities, competitive preferences, scoring criteria, detailed evaluation and management plans, and a frustrating online submission system were all part of the process.

While these requirements should be taken seriously, the underlying criteria were much simpler. The i3 application process was about finding organizations with: (a) a good idea for improving education outcomes for high-need students in K-12 schools, preferably an idea with at least some evidence or a solid theory behind it; and (b) the capacity to implement, evaluate, and scale that idea effectively.

What helped the grantees win? When asked, they had the following suggestions for future applicants:

- **Meet One of the Department's Chosen Absolute Priorities:** To qualify, the applicant had to apply in one of the Department of Education's chosen focus areas, called absolute priorities. The topics differed from year to year, but in 2010 they were the four categories discussed in the previous chapter: (1) teachers and principals; (2) data use; (3) standards and assessments; and (4) school turnarounds.

In later years the Department chose other topics, such as STEM, school climate, family and community engagement, students with disabilities, English language learners, technology, non-cognitive skills, and rural

communities. There were also "competitive priorities," which could provide added points on a grant application. Examples in the first year included early learning, college access, student with disabilities, and rural communities.

From the Department's perspective, this encouraged applicants to apply in research areas where there was a greater need for more evidence. It also created cohorts of grantees with programs that were similar enough to allow cross-project learning and communities of practice.⁵⁶

In interviews, some of the grantees believed the revolving categories made it impossible to know in advance if they would be eligible to apply for a new award when their current grant was complete. This could make sustainability planning more difficult. On the other hand, some of i3's priorities were broad enough that they allowed nearly any project to apply so long as they included one or more of the required components (such as professional development or data use). The large number of reading and literacy-related grants in the first year's cohort, which were evenly distributed across each of that year's four absolute priorities, was an illustration of how much flexibility there was.

- **Have a Good Idea:** The grant application relied on technical evaluation jargon to describe its requirements, but the lower-tier development grantees saw it in much simpler terms: you need the right people and a good idea, particularly one that was consistent with the organization's overall mission.

"Evidence-based funding is important, but I am skeptical of some of what passes for education research," said one grantee. "The conditions are hard to replicate and transfer. You need good ideas, good people, and the latitude to make it work."

"Don't apply to do something brand new that you

⁵⁶ Interview with OII staff, September 15, 2016.

can't sustain when the money is over. Apply to do more of what you are already doing and that you want to study," said another. "Make sure what drives you are the questions, not the money to fund the program. No grant truly pays for itself," said another.

"We have been around for 30 years. We are constantly innovating practices. Continuous innovation cycles define who we are as an organization," said another grantee. "We didn't invent this intervention for i3. We had 20 years of program development under our belt already. What i3 allowed was for us to distil and pull out those most promising elements and apply for an R&D grant."

- **Meet the Evidence Requirements:** While good ideas were necessary, a proposal's evidence still mattered. Better evidence also seemed to improve a grantee's prospects for success (as discussed in the last chapter).

In the 2010 competition, the large scale-up grants had to be supported by strong evidence (usually at least one large or two smaller RCT-based studies or equivalently rigorous evaluations). Validation grants needed moderate evidence (at least one well-designed and well-implemented impact study), while the smallest development grants only needed a reasonable hypothesis.

After the applications were submitted, staff from the Institute of Education Sciences (IES) would review the citations in the application to determine if they met the evidence standards. If one did not, the Department ruled it ineligible and dropped it from the competition.⁵⁷

"What positioned us was an RCT study in the past that met What Works Clearinghouse standards. Also, we had the STEM focus," said one grantee. "We were not doing anything brand new, but we were combining pieces that have evidence in new ways," said another.

- **Hire an Experienced Evaluator:** Applicants also needed to include an initial evaluation plan. "When you apply for an i3 grant, the evaluator writes that part," said one grantee. "If they don't have i3 experience they will be surprised by the evaluation design requirements. They will have people come in and make sure the design meets

the What Works Clearinghouse expectations."

The importance of choosing the right evaluator cannot be overstated. This decision could, by itself, determine whether a project would succeed or fail, creating reputational risk for project partners. More information about evaluator qualifications and evaluation generally can be found in Chapter Four.

In choosing an evaluator, projects may wish to pay attention to intellectual property rights to determine who may publish results from the project.

- **Have Experience with the Proposed Program:** "We proposed something that we already knew how to do, wanted to do it in more schools, and to study it," said one grantee. "A lot of folks were proposing something that was evidence-based, but they had not done it themselves."
- **Ensure Genuine Commitment from Partner Schools:** Nearly every i3 project takes place in the schools. While the i3 application required evidence of partner commitments, however, most applicants could satisfy them with pro-forma letters of support and memoranda of understanding (MOUs). Several grantees suggested that this would be a mistake, however, and if it happened it could be a disaster for the project. Poor school buy-in usually produced poor implementation and poor evaluation results.
- **Have Match Commitments Ready:** Winning applications were required to provide a private match, usually from foundations or corporate partners. Donations could be cash or in-kind. While the program allowed applicants to wait until after they won the grant to meet this requirement, it was frequently a highly stressful moment for those who did not have solid initial commitments in place. In later years, the match requirements were lower, but so too was the interest of national funders in i3 projects, according to several grantees who applied in those years.

This issue may have been partly addressed under the new EIR program, however. In addition to the private support allowed under i3, it also allows support from other federal programs, states, and local governments to be counted toward its 10 percent match commitment.⁵⁸

⁵⁷ The timing of this evidence review varied from year to year. In the 2010 application process, it came after independent peer reviewers scored the applications.

⁵⁸ See Section 4611(d) of the Elementary and Secondary Education Act as amended by ESSA. Available at: <http://socialinnovationcenter.org/wp->

- **Have a Realistic Scaling Plan:** Plans for taking an initiative to scale were an obvious requirement for the scale-up grants, but the validation and demonstration grantees also usually included plans for expanding into new schools. Among other benefits, working with enough schools or students was necessary to generate a sample large enough to meet the project's evaluation requirements. However, some grantees seemed to overpromise, which could stretch them too thin.
- **Hire an Experienced Grant Writer:** A talented grant writer usually made a big difference, particularly one who understood how to maximize points from the peer reviewers. "We hired a professional grant writer with experience with federal grants," said one grantee. "We focused on what we knew best. Then we had one internal person write it with one voice," said another.
- **Be Persistent:** Some who lost in the early years tried again later and won. "We took the reviewers' comments seriously and studied the successful proposals that were posted," said one grantee.

"We applied twice and were not successful. The third time was the charm," said another. "We are a nonprofit and not a district. We learned a lot from the comments we got back. By the time we applied the third time a lot of our folks said don't bother, you won't get it, but we did."

Program Launch

For most of the grantees, receiving word that they had been selected produced an explosion of activity, including fundraising, hiring, training, confirming partnership agreements, and solidifying evaluation plans.

Grantees that were building on pre-existing work appeared to have the easiest time during the launch period. This experience seems consistent with those in another federal tiered-evidence program, the Social Innovation Fund, where grantees launching new programs with new staff in new settings faced the greatest hurdles, while those that expanded existing services were more likely to succeed.⁵⁹

"What we applied to do we had already started doing with a more modest grant," said one grantee. "The work was already happening, so it was just a little extra love," said another.

For many others, it took more work. Some built planning periods into their grant applications. "We had six months of planning," said one grantee. "We had a development year and a pilot year," said another.

Others launched almost immediately. "The smart people, not us, put time for themselves in the beginning. We didn't have that," said one grantee. "We were notified in November, the launch was in January," said another.

The timing of the grant announcements varied from year to year, and sometimes they did not mesh well with the school calendar. "Timing with respect to the school year made things difficult in that first year. It made that year not as strong as it would have been," said one. "Our school district partners have been great, but schools are in turmoil at the beginning of the school year," said another.

In interviews, the grantees cited the following as the most typical launch period activities:⁶⁰

- **Finding Match Funders:** One of the most pressing initial tasks for new grantees was finding their match funding. According to the Department of Education, none of the i3 grant winners have failed to obtain their match, but it was not easy.

Some thought the match requirements were helpful. "Being able to say we had a 5-1 match was attractive to funders. We made connections we wouldn't have made," said one grantee. Others disagreed. "Don't require innovation grants to come up with private funding. It stifles. You are cutting people off who have ideas and don't have connections with the funding community," said another.

In some cases, the timing of the announcement made things more difficult. "We had to rally in August, which is a challenging time for grant funders," said one grantee.

To help i3 applicants, the Department of Education recruited twelve of the nation's biggest

⁵⁹ [content/uploads/2015/12/ESSA-EIR_Provisions.pdf](http://www.socialinnovationcenter.org/wp-content/uploads/2015/12/ESSA-EIR_Provisions.pdf)
Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, pp. 17-18. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-

⁶⁰ [30.pdf](#)
This section is based on project director answers to the following open-ended interview question: "Were there any challenges associated with the initial launch of the initiative?"

foundations to provide \$500 million in matching funds.⁶¹ Applicants could submit proposals through a common website, called the i3 Foundation Registry.⁶² While the registry still exists, however, interest from the foundation community appears to have declined. "The i3 Registry didn't pan out for us. We never got any hits or inquiries," said one grantee.

Over time, the match requirements were reduced, but some of the grantees thought the fundraising environment had also become more difficult. "It was the Wild West back then, but it might have been easier to get that money then than now," said another. "They knew what the goal was and that ED was going to send them the cream of the crop. Now it seems like you are kind of on your own. You make an approach and you may get crickets."

As noted elsewhere, the match requirements have been changed under EIR. In addition to private support, the program also allows grantees to count cash or in-kind support from other federal, state, or local government sources.⁶³

- **Onboarding New Staff:** Another major early focus was hiring and training new staff. The project director was often one of the first to be hired, and along with the choice of evaluators was probably the most critical hiring decision. Staff turnover was also a regular concern. "Be prepared for personnel changes. That will happen. It is going to be a wild ride," said one grantee.
- **Recruiting Schools:** While many projects specified partners in their applications, a lot of recruitment happened during the ramp-up period. "The big thing to recognize is that the commitments going in don't make any difference after you got the grant. Principals change," said one grantee. "Principals are not necessarily bought in," said another. "Districts would say yes and then tell the principals."

Orientations for the schools were common. "It's about hitting the road. Signing them up and developing the MOUs," said one grantee. One sticking point with the schools for several

grantees, however, was the use of RCTs. "We did a RCT. It took a little convincing to get the local programs to buy into that," said one.

- **Planning and Preparing Evaluations:** Local evaluators were also busy during this period. They were required to submit formal evaluation plans to the Department of Education's technical assistance coordinator, Abt Associates, and sometimes that led to surprises.

"We didn't know the role of Abt or that they could change things as much as they did," said one grantee. "We had initial discussions with school districts based on our application. Abt said we had to do an RCT, which we had not planned for."⁶⁴

Some of the grantees felt pressure to implement their programs with fidelity early on. "On program fidelity, it was okay but it got better," said one. "We had to have fidelity on day two," said another, although in this case it was internal pressure because the project's leaders wanted to show impact during what was a relatively brief study period.

Other early-stage evaluation activities included obtaining necessary Institutional Review Board (IRB) approvals, setting up data systems, and working out final data agreements with the schools.

- **Establishing Grant Management Infrastructures:** Some grantees that lacked experience with federal grant requirements were surprised by the burden. "Federal grants have lots of compliance issues. You need an infrastructure for a large federal grant," said one. "We hired our first CFO. We had a controller and accounting already, but the CFO was great."

School Partnerships

Partnerships played an important role in every i3 project. These partners included schools, colleges, businesses, funders, and technology providers, among others.⁶⁵ Of these, the relationships with the

⁶¹ Ron Haskins and Greg Margolis, *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*, Brookings Institution Press, 2015, pp 117-118.

⁶² See <https://www.foundationregistryi3.org/>.

⁶³ See Section 4611(d) of the Elementary and Secondary Education Act as amended by ESSA. Available at:

http://socialinnovationcenter.org/wp-content/uploads/2015/12/ESSA-EIR_Provisions.pdf

⁶⁴ This was actually a negotiation process. For more details, see the section on evaluation.

⁶⁵ This section is based on project director answers to the following open-ended interview question: "Were there any

schools were usually the most important. Nearly all of the i3 projects were conducted in schools and genuine buy-in and capacity in those schools were usually major determinants of a program's success.

School cooperation and capacity affected a wide range of issues, including willingness to implement a program with fidelity and data access. Projects that were unable to convince teachers or other critical school personnel to fully buy into a tested intervention were often doomed from the start.

School Buy-in

Genuine buy-in from the schools was critically important to project success. "You want a stable and supportive school district that is willing to make changes when changes are requested," said one grantee.

"The odds of making major change are low if you don't bring in advocates among the teachers and others on the ground," said another. "Innovations are too easily blocked otherwise. It is like white blood cells in an immune system reaction. They can come at you in any number of ways. You should know that going in."

Some initiatives benefitted because their project leaders were located in the schools. "Because we are in the school, it makes things easier," said one grantee. "The nonprofits at i3 conferences envied our position at the schools," said another. "We did not get significant kickback from schools or principals. There were benefits from that standpoint. We had more leverage."

Such buy-in could not be assumed. Many schools are used to working with outside partners and their commitments could be pro-forma. Often overwhelmed, they sometimes see such outside projects a source of extra resources and were not used to making commitments. For i3, sometimes school leadership could even not remember the commitment when the award had been made. "When we inform the schools and told people they won, there could be school amnesia," said one grantee. "We won what?"

"We needed to prove ourselves in the schools," said another grantee. "We needed the administration and teachers to know we weren't coming and then vanishing. We were there for the long haul."

Gaining approval from just one level of

management was usually not enough. "Some principals don't involve others in decision making. You need to know the leadership culture and how they interact with district leadership. Some school partners never look at the MOUs. Others do and are attentive," said another grantee.

"When we were writing the grant, we were very focused on district buy-in and community buy-in," said one grantee. "We weren't as focused on teacher buy-in, which is hard when you are just writing a proposal. Had we involved more teachers in the very beginning, we might have had a smoother time getting them excited about it."

"We get staff to vote," said another grantee. "We need 80 percent of the teaching staff to vote in favor of doing this. We are not imposing it on people."

School Capacity

School buy-in was only half the equation. The other half was having enough capacity to act on that buy-in. In many schools, particularly those that were low-performing, this could be a major challenge.

"Most of our schools were turnarounds," said one grantee. "Their principals can barely breathe. Unless you can really contribute to meeting their needs, you aren't in a real partnership. You aren't really helping them. They will only help you to the ability they are able, which isn't very much."

"Turnarounds are painstaking work and i3 just wasn't a focus for them. In reality, this was just one part of many things for those schools," said another grantee.

"Our principals were compelled to accept a 16 percent pay cut to their contract. Teachers' benefits were renegotiated—to their disadvantage," wrote another in a project where several schools were closed.

Ongoing staff turnover could also play a major role. "One of the major things we learned is that it is difficult to work with school districts where people keep changing all the time. We get MOUs and then the new principal comes in and doesn't know what the old one agreed to do," said one grantee.

"We got the i3 grant during a period of turbulence in the teacher hiring landscape due to severe budget constraints. Having partnerships with a diversity of schools allowed us to persist," said another grantee. "One school district riffed all of its teachers with less

than three years of seniority. So our teachers lost their jobs. We were able to replace those teachers and the next group of people because we had relationships with other schools and districts. Those deep partnerships at the regional level, keeping in close touch, that is absolutely crucial.”

More information on recruiting and working with the schools can be found in the Scale section of Chapter Four.

Capacity Building

Successfully implementing evidence-building grants like i3 requires substantial capacity. The success of grantees in another tiered-evidence initiative launched at about the same time, the Social Innovation Fund, also depended greatly on the capacity of its grantees.

The most successful projects in that program were well-resourced, featuring strong leadership and organizational cultures, deep experience with evidence-based programs, financial management and compliance systems, performance management systems, fundraising ability, and substantial evaluation capacity. By contrast, poorly-resourced projects usually struggled.⁶⁶ As of June, 2015, none of these projects had produced a rigorous impact evaluation with positive results and several had dropped out of the program.⁶⁷

This pattern of higher capacity organizations producing better results appears to be repeating in i3, with grantees in the higher tier scale-up and validation grants more likely to generate positive evaluation findings than those in the lower tier development grants (although higher levels of incoming evidence have probably also played an important role). The variable capacity of the grantees could also be seen in the performance reports,⁶⁸ interim implementation evaluations, and progress on fidelity metrics.⁶⁹

Unsurprisingly, capacity building has been a central focus of the i3 program. While the program has provided substantial technical assistance, however, so far it has met with mixed results.⁷⁰

Program Officers

The primary task of overseeing the grantees fell to program officers housed at the Department of Education. Program officers monitored project performance and compliance with regulatory and budgeting requirements through reports and regular phone calls. They reviewed and approved any proposed project changes and sometimes provided technical assistance.

When performance and capacity issues arose, the local project directors were the first to know. This information was relayed back to the i3 program officers and, secondarily, through check-ins with the two technical assistance providers (Westat and Abt Associates), which are discussed below.⁷¹

The program officers were generally well-regarded by the grantees. However, there was less appreciation for the paperwork requirements and some felt that turnover in the positions hurt the program, forcing local project directors to reexplain their initiatives to new staff who did not know their history.

The Department assigned a relatively low 10-12 grantees to every program officer, which seemed to make a difference. “It was a high touch, strong relationship. I’m not used to that, but it helped,” said one grantee. “It keeps me on track and keeps the project on track,” said another. “It’s not possible for me to drift very far because I need to respond to them.”

Technical Assistance Providers

For further assistance, the Department contracted with two external technical assistance providers, Westat as the primary TA provider and Abt Associates for help with evaluations.

Abt was hired in September 2010 and it appears to be well regarded by the grantees. It worked primarily with evaluators and only a few of them participated in the project director interviews, but of the nine groups that provided a judgment on its technical assistance, eight rated it positively. Abt’s

⁶⁶ Social Innovation Research Center, “Social Innovation Fund: Early Results Are Promising,” June 20, 2015, pp. 36-37. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf

⁶⁷ Ibid., pp. 16-17

⁶⁸ Obtained through a Freedom of Information Act (FOIA) request, as discussed in the methodology section.

⁶⁹ The central role of capacity in determining overall project

success is discussed in the Reasons for Success or Failure section of Chapter One.

⁷⁰ This section is based on project director answers to the following open-ended interview question: “What were your critical technical assistance and capacity needs? Did the program’s TA providers address these fully?”

⁷¹ The level of communication between the program officers, Abt, and Westat was unclear. No program officers or representatives of Westat were interviewed for this report.

role is discussed in greater detail later in the report.

The Department contracted with Westat in 2012. After conducting a needs assessment to identify high priority grantee needs, it ran a series of webinars, launched an online learning community,⁷² and organized communities of practice that brought grantees together on issues of common interest, such as STEM, school turnarounds, rural issues, sustainability, and scaling.⁷³

“The communities of practice worked,” said one grantee. “I learned from every one that I participated in. It took a long time, maybe not until year five was it really working, but we would like to keep participating. That was a success.”

“Use the i3 community, but don’t limit yourself to the TA structures,” recommended another. “Reach out to the other i3 grantees. Those informal conversations are really invaluable. They want to keep track of what impact they are having, but some of it is organic.”

“What was helpful was when they did some initial presentations on the regulations,” said one grantee. “There were a lot of support materials. I watched the recorded webinar several times.”

The grantees also liked the annual conferences for project directors, which were a joint effort between Westat, Abt, and the i3 program staff. “They do a good job with the annual conference,” said one. “The conferences got better every year,” said another.

Some grantees, however, wanted more tailored assistance. “It needs to be more customized so the guidance is specific to our needs, not generic advice,” said one grantee. “What we really needed was a full-time consultant,” said another.

Westat has provided some targeted assistance, but it has been focused on grantees with clearly identified needs that were a good match with the resources that Westat either had available or could customize and leverage. Examples have included literature reviews, assistance with logic models, help with sustainability plans, and site visits.⁷⁴

In interviews, several of the grantees expressed a strong preference for more individualized assistance of this kind, particularly from acknowledged national experts in their areas of interest. “Make the pool of experts wider. Make it

more organic and interactive with the field at large,” said one grantee.

William T. Grant / Spencer Foundation i3 Learning Community

One independent effort that was very highly regarded by the grantees who were invited to participate was the i3 Learning Community sponsored by the William T. Grant and Spencer Foundations. This initiative, which lasted from 2011-2014, brought together 17-20 grantees twice per year to share insights and challenges.⁷⁵ “That interaction was perhaps the most beneficial part of the whole process,” said one grantee that participated.

What set it apart? “The national gathering of the project directors, no matter what ED says, there is a feeling of compliance about it,” said one grantee. “Some of the sessions are helpful, but it is by the book. It’s very much what you would expect from a federal agency. Abt’s role was to be the TA provider, but whenever we met with them we felt like we were meeting with the teacher. We had to do it right.”

“This meeting, from the beginning, felt like it was our group. It was collaborative. We shaped the agenda. They had hired a group to provide TA to the learning community and they were 100 percent responsive. They would call and ask us what we wanted. If you could have been a resource you were asked to lead a session. It was a smaller group too – maybe three people from each project. So you had 50 people that you see twelve times rather than 200+ once a year. There was no expectation but to learn. It was delightful.”

Internal Capacity Building by the Grantees

Most grantee capacity needs were met internally. In interviews, some said their capacity was already sufficient, or at least sufficient enough that they did not need help from the Department’s TA providers. “We are a big organization and have a lot of expertise,” said one. “We have a whole research team. If we were smaller, we would need the help.”

For others, a large amount of capacity building came from the grant itself. The importance of this to the scale-up grantees, who received grants ranging from \$45-50 million, is described in Chapter Four, but

⁷² See <http://i3community.ed.gov>

⁷³ Communication with the Office of Innovation and Improvement, January 12, 2017.

⁷⁴ Ibid.

⁷⁵ Resources from these meetings are available on the Forum for Youth Investment web site at:

<http://forumfyi.org/WTGrantFoundationi3>

this was also true for the smaller grantees. “We had never gotten a federal grant before, so we had never gone through an audit that was associated with a federal grant,” said one. “That process helped us develop better back office procedures. We needed to do that to grow. There wasn’t a lot of internal will to do that. We were forced to do it and it was good for us.”

Where they had specific needs, many grantees addressed them on their own. “We have taken charge of our own capacity building,” said one grantee. “We went outside the i3 community and created a school change advisory committee. We pulled together a set of experts to get their feedback. That advisory group was written in as part of the grant.”

Finally, the needs of the i3 grantees were distinct from those of the schools they were working with, many of which experienced substantial capacity

shortfalls (for more, see the previous section on School Partnerships). For most of the grantees, addressing these needs was a central component of their work. Training and other capacity building for local schools were common program components. Most of them handled this on their own.

“We do a lot to build capacity. There is district level coaching, school level, and workshops and conferences,” said one. “That capacity building is the heart of our project.”

There was significant evidence, particularly among the scale-up grantees, but also across most of the projects, that many of the grantees were playing centralized intermediary or backbone functions.⁷⁶ Their success in this role was a major determinant of the project’s overall success. This seemed especially true in projects working with a large number of low-capacity schools.

⁷⁶ See the Scale section of Chapter Four and the Reasons for

Success section of Chapter One.

Chapter Three: Innovate

What did it take to successfully execute an i3 project? This chapter reviews several related issues, including: (1) program development and innovation; (2) evaluation; and (3) obtaining needed data.

Innovation

The Investing in Innovation program was designed to "create an innovation pipeline" in K-12 education, according to Jim Shelton, a former deputy secretary at the Department of Education who oversaw it in its early years.⁷⁷ When the first round of grants was announced in 2010, however, the program drew fire from some prominent education experts for the "disappointing and 'been there, done that' nature of so many 'innovative' winners."⁷⁸

Most of the criticism was directed toward the comparably well-known scale-up grantees like KIPP, Teach for America, Success For All, and the Reading Recovery initiative overseen by Ohio State University. The big scale-up grants were always intended to go to programs with strong evidence and solid track records, however.

Less attention was paid to the lower-level grants, particularly the development grants, where many comparably unknown grantees had won smaller awards for relatively untested innovations. At the time, it was not obvious how new, groundbreaking, or successful these grants would be.

As final evaluation results have begun to appear, this has become more clear. As might have been expected, there have been several successes among this group and those grants are described elsewhere in this report (including Appendix A). Other grants produced null findings, but this was also expected.⁷⁹

The overall failure rate has not been substantially different from what is commonly experienced in the private sector.⁸⁰

These summary results only scratch the surface, however. One question is whether the successful development grantees have produced any innovations that are truly groundbreaking. This is a subjective question that is touched on in Chapter One,⁸¹ but the best that can probably be said is that the program's progress so far represents a good start, but there is substantial room for improvement.

Interviews with project directors, coupled with a review of their evaluations and internal performance reports, suggest that the program could be better designed to select, support, and evaluate truly innovative initiatives.

Choice of Interventions

Successful innovation begins with a good idea, ideally one that is well-grounded in research and represents a promising and potentially groundbreaking improvement on current practice. While the i3 program has sought such projects, however, it is not clear how effective it has been at finding them.

The application requirements for i3 development grants varied from year to year, but the application in 2010 (the year most relevant for projects that have produced final evaluations so far) defined what it was seeking this way:

Development grants provide funding to support high-potential and relatively untested practices, strategies, or programs whose efficacy should be systematically studied. An applicant must provide

⁷⁷ Alyson Klein and Sarah D. Sparks, "Investing in Innovation: An Introduction to i3," *Education Week*, March 22, 2016. Available at: <http://www.edweek.org/ew/articles/2016/03/23/investing-in-innovation-an-introduction-to-i3.html>

⁷⁸ Rick Hess, "i3 Winners: Long on Talent, Execution, & 'Best Practices'—Not Transformation," *Education Week*, August 6, 2010. Available at: http://blogs.edweek.org/edweek/rick_hess_straight_up/2010/08/i3_winners_long_on_talent_execution_best_practices—not_transformation.html

⁷⁹ Of the 30 development grants that have released final evaluations so far, 16 produced no impact and 6 others only

reported basic information on outcomes. Six reported positive impacts and another two produced mixed results. These results are summarized in Chapter 1 and described in greater detail in Appendix A.

⁸⁰ Studies of new products or strategies by for-profit organizations like Google or Capital One routinely experience failure rates of 90 percent or more. See Jim Manzi, *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*, Basic Books, 2012, pp. 143-167. Similar failure rates are common in health research and business formation.

⁸¹ This is discussed in the Progress in Priority Areas section.

evidence that the proposed practice, strategy, or program, or one similar to it, has been attempted previously, albeit on a limited scale or in a limited setting, and yielded promising results that suggest that more formal and systematic study is warranted.⁸²

The process of finding such innovations began with the applicants, who were required to describe their interventions in their proposals. In interviews, project directors said that these early design decisions were typically made in a collaborative process that involved the grant writer, relevant organizational personnel, and the evaluator.⁸³

After the applications were submitted, the Department's selection process varied from year to year, but it usually involved some combination of reviews by: (1) OII staff, who confirmed applicant eligibility and oversaw the process; (2) staff at IES, who reviewed the evidence cited in the applications; and (3) external peer reviewers, who scored the applications according to criteria that included the project's significance, personnel qualifications, organizational capacity, and management, evaluation, and scaling plans.

From 2010-2015, the Department received almost 5,000 applications, most of which were in the development grant category. With an overall selection rate of about 3 percent (lower than that of Harvard, Stanford, or other highly competitive colleges),⁸⁴ the program seemed well positioned to find the cream of a large crop.⁸⁵

Whether it was successful is debatable. The Obama administration's competitive grantmaking process, particularly its peer review process, has been criticized for several shortcomings. One review of competitive evidence-based grant programs,

including i3, found that their peer review processes often suffered because they:⁸⁶

- Often disallowed the most informed reviewers because of perceived conflicts of interest;
- Typically did not include an in-person meeting with the applicant's leadership team; and
- Typically did not allow reviewers to examine external information to assess claims made in an application.

All three were true of i3.⁸⁷ According to *Education Week*, the i3 process also suffered because different peer reviewers used different rating systems, which produced scoring anomalies.⁸⁸ There was also evidence of that in the interviews for this report.

"I have found that the feedback and scoring are all over the place," said one grantee. "Some of the comments are related to criteria that are not there or they are factually incorrect. A reviewer might say there is no research basis even when we cite something. You can't fix the errors they make."

This experience is also consistent with that in the Promise Neighborhoods program, another competitive grant program run by OII, where peer reviewers also reported a substantial degree of subjectivity in their scoring.⁸⁹

Former Obama administration appointees have acknowledged that choosing the most promising development grantees is a challenging undertaking for any government program, including i3.⁹⁰ One suggestion is that development grants might be improved by utilizing a two-tier process similar to that used by the Social Innovation Fund, where grantees are chosen by intermediaries after thorough vetting and due diligence.⁹¹

⁸² Information about the 2010 competition can be found at : <https://www2.ed.gov/programs/innovation/2010/applicant.html>

⁸³ This question was explored in project director interviews with the following open ended-question: "What strategies did you use for obtaining the grant? What do you think made it stand out?" A review of the answers to this question can be found in Chapter One.

⁸⁴ U.S. News and World Report, "Top 100 – Lowest Acceptance Rates." See: <http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/lowest-acceptance-rate>

⁸⁵ The Department received almost 5,000 i3 applications or pre-applications between 2010 and 2015, but made only 156 grants, for a total application-success rate of 3.1 percent. See U.S. Department of Education, "Innovation and Improvement: Fiscal Year 2017 Budget Request," p. F-31. Available <https://www2.ed.gov/about/overview/budget/budget17/justifications/f-ii.pdf>

⁸⁶ Shivam Mallick Shah and Michele Jolin, "Social Sector

Innovation Funds: Lessons Learned and Recommendations," Center for American Progress, November 2012, pp. 24-27.

Available at: <https://www.americanprogress.org/wp-content/uploads/2012/11/JolinInnovationFunds-2.pdf>

⁸⁷ Interview with OII staff, January 17, 2017.

⁸⁸ Alyson Klein and Sarah D. Sparks, "Investing in Innovation: An Introduction to i3," *Education Week*, March 22, 2016. Available at: <http://www.edweek.org/ew/articles/2016/03/23/investing-in-innovation-an-introduction-to-i3.html>

⁸⁹ Patrick Lester, "View from the Inside: The Promise Neighborhoods Peer Review Process," Social Innovation Research Center, October 10, 2010. Available at: <http://www.socialinnovationcenter.org/?p=2367>

⁹⁰ Several former Obama administration officials discuss this on the record in the Epilogue.

⁹¹ Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, p.p. 19, 34-38. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-

Support for Continuous Improvement

The literature suggests that successful innovation relies on a firm understanding of the existing research, a strong grounding in theory and logic models, and an iterative process of experimentation, studying results, and continuous improvement.⁹²

In theory, this process was open to the grantees. While the original project designs described in the grant application served as an overall framework, the grantees could request changes subject to the approval of the i3 program officers. Some did.

In practice, however, the flexibility to engage in iterative change and continuous improvement may have been substantially limited in many (and perhaps most) cases. The most critical actors at this stage were the local project directors, who had nominal authority over their projects. For many, however, the authority to make improvements appeared to be inhibited by several factors.

One was that they were often hired after the grant was received and had no input on the original program design. Some may have believed that their responsibility was to implement the program as designed. Others who wished to make adjustments may have faced real or perceived barriers, not just from i3 program officers whose approval was needed, but also from constraints imposed by their evaluations (particularly the need to comply with fidelity measures), needed permissions from internal management, and flexibility from external partners, including the schools.⁹³

Another potential limitation was that, as a group, many project directors appeared to have an uneven understanding of the research in their respective fields of interest.⁹⁴ For some, this may have been overcome because they were working within large, high-capacity organizations, many of which had separate research offices or other staff with extensive experience with the tested intervention.

Other grantees were comparably lower-capacity nonprofits or local school districts. Many of them appeared to be managing projects with little external support. In these cases, the project directors might have benefitted from a closer relationship with national experts in their respective fields. Several said so in the interviews.⁹⁵

Some of these factors may be changing, however. The program's latest competition, the first issued under the newly reworked and renamed EIR program, appears to be encouraging continuous improvement in its lowest tier. Its new version of a development grant, called an "early-phase" grant, now includes the following description:⁹⁶

The first years of an Early-phase grant are expected to focus on developing and iterating the practice in a few schools (or a limited version of the practice in a greater number of schools), and the independent evaluation is expected to generate information to inform the practice's development and iteration; the remaining years of an Early-phase grant are expected to entail full-scale implementation across the project's full set of schools,

While this new flexibility is a step in the right direction, earlier experiences from i3 suggest that more support will be needed for grantees to make it work effectively.

Appropriate Evaluation Methods

Finally, several of the development grantees working with early-stage, developmental interventions may have subjected them to high-level evaluation methodologies, such as randomized-controlled trials, before they were ready.⁹⁷ Evaluation is described in more detail in the next section.

In interviews, several grantees offered suggestions for further improving the program's approach to innovation more broadly:

⁹² [30.pdf](#)
Examples include the Deming Plan-Do-Study-Act cycle and rapid-cycle improvement. See <https://en.wikipedia.org/wiki/PDCA> and Mathematica Policy Research, "Rapid-Cycle Evaluation" at <https://www.mathematica-mpr.com/our-capabilities/rapid-cycle-evaluation>

⁹³ The ability to get partner schools to go along with requested changes came up frequently in the interviews.

⁹⁴ This conclusion is based on answers to four open-ended interview questions about the project's research base, one of which was: "In your opinion, how strong is the existing evidence base in this focus area? What do we know and what

don't we know?"

⁹⁵ This is described in greater detail in the previous section on capacity building.

⁹⁶ See *Federal Register* notice at <https://www.federalregister.gov/documents/2016/12/15/2016-30085/applications-for-new-awards-education-innovation-and-research-program-early-phase-grants>

⁹⁷ The process for reviewing and approving evaluations is described in the next section. In general, the final evaluation was the result of negotiation between Abt, the federal project officer, and the local grantee. While project officers pressured grantees to meet high standards, they lacked authority to force a grantee to adopt a certain methodology if they resisted.

- **Greater Use of Pilot Phases and Formative Evaluations:** Some of the development grantees created a learning environment for themselves with pilot phases and relied more on formative or interim evaluations, which allowed for adjustments and improvements before conducting their final impact study. For example, the American Federation of Teachers conducted a small-scale pilot of its teacher evaluation system during the 2010-2011 school year before rolling it out more broadly.⁹⁸ Similarly, the California Education Roundtable spent almost two years developing, renewing, and piloting its project-based math intervention before it began its randomized controlled trial.⁹⁹

- **Greater Use of Rapid-cycle Testing and Continuous Improvement:** “We need to find ways to improve programs, practices, and systems,” said one grantee. “Let’s not be too hasty in abandoning approaches that do not instantly pay off. After all, many established interventions had years to gestate. Let’s not cut short this process for new innovations that are just starting out.”

“Make the development grants more innovative and more flexible,” said another. “Make it more about continuous learning and testing – more of an R&D process.”¹⁰⁰

“Think about ways of incorporating more rapid improvement cycles so it is not all reliant on the summative assessment at the end of the year,” argued one. “Getting and sustaining good outcomes is about a hundred different one percent solutions, not one silver bullet,” said another.

- **Greater Use of Single-School Projects:** Studies in single schools may be appropriate for

early-stage projects, allowing greater focus and the flexibility to test new innovations. The Bellevue School District’s career and college readiness project, which was conducted in a single school, was an example.

Studies of such projects will face generalizability barriers, however, and they will not receive a high rating from the What Works Clearinghouse.¹⁰¹ Under EIR, they are no longer allowed,¹⁰² but this was an administration decision, not one required by law.

- **Less Focus on Fidelity in the Early Stages of the Grant:** “Fidelity is less useful for development grants,” said one grantee. “When you are trying to do something innovative, trying to do fidelity presupposes you have figured everything out.”
- **Greater Use of Qualitative Evidence:** Several of the grantees also incorporated substantial qualitative feedback into their studies. The evaluation of the Bellevue School District’s career and college readiness project, which included extensive qualitative and exploratory analysis for an intervention in a single school, was a good example.¹⁰³
- **More Testing of Existing Innovations:** One criticism of i3 and many of the other Obama administration innovation programs¹⁰⁴ is that five years is a long time to wait for results. One of the major reasons for this lengthy delay is the time needed to put new programs in place. Many promising programs are already operating and fully funded, however. For these projects, a comparably small amount of funding could provide an evaluation component (and possibly randomization) that would provide comparably rapid, inexpensive, and rigorous testing for

⁹⁸ AIR, “The Educator Evaluation for Excellence in Teaching and Learning (E3TL) Consortium Evaluation Report,” September 2015. Available at: https://i3community.ed.gov/system/files/resource_files/2016/e3tl_evaluation_final_report.pdf

⁹⁹ WestEd, “STEM Learning Opportunities Providing Equity: An Investing in Innovation (i3) Grant Final Evaluation Report,” September 30, 2015. Available at: http://arches-cal.org/wp-content/uploads/2016/06/WestEd_Final-Evaluation-Report_SLOPE-i3_09-30-2015.pdf

¹⁰⁰ For example, see Mathematica Policy Research, Rapid-Cycle Evaluation: A Primer, February 2016. Available at: <http://www.mathematica-mpr.com/our-capabilities/rapid-cycle-evaluation>

¹⁰¹ Single-school studies present possible confounding problems that undermine claims of impact under WWC standards. For more on confounding, see Andrea Skelly, et al, “Assessing Bias: The Importance of Considering Confounding,” Evidence-

based Spine Care Journal, 2012. Available at:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3503514/>

¹⁰² The Department of Education’s policy under EIR is that early-phase projects must be implemented in multiple sites to increase generalizability and resolve potential confound issues.

¹⁰³ Randy Knuth, Ph.D., et al, “Re-imagining Career and College Readiness: STEM, Rigor, and Equity in a Comprehensive High School,” March 1, 2016. Available at: <http://www.bsd405.org/wp-content/uploads/2016/03/Sammamish-i3-Grant-Findings-Report.pdf>

¹⁰⁴ This has also been true for the Social Innovation Fund. See Social Innovation Research Center, “Social Innovation Fund: Early Results Are Promising,” June 20, 2015, pp. 20-21, 30-38. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf

innovations that are already underway.

- **Low-cost, Short-duration Grants:** Another possible option is to allow lower cost, short-duration grants like those that have been funded by the Institute of Education Sciences.¹⁰⁵ The first of these grants were announced in 2016 and they may be a model for i3 (now EIR).¹⁰⁶
- **Greater Tolerance for “Failing Forward”:** Truly new and innovative interventions are more likely to fail. Some grantees thought that there should be lower expectations for development grants. “It is hard to convince people to innovate in a punitive situation,” said one grantee. “People can lose their jobs. They don’t feel free to try something and fail.”

“Good innovation sometimes has failures – sometimes big ones. There doesn’t need to be negative accountability,” said another. “But if there is early data that things are not going the right way then there should be parameters so we can pull the plug.”

Evaluation

A central component of every i3 project is its independent evaluation. How did the grantees develop their evaluations?¹⁰⁷ What factors contributed to their success or failure?¹⁰⁸

This section reviews: (1) grantee experiences with their program evaluators; (2) the federal role; (3) effective program implementation and model fidelity; and (4) grantee recommendations for achieving successful evaluation results.

The Evaluator Role

Evaluation is a central component of the i3 program and independent evaluators, who interfaced

with both local project personnel and the national technical assistance team at Abt Associates, were central to that process. For many i3 projects, the choice of evaluators was a major determinant of their success.

How did they choose their evaluators? For most projects, this decision came early, before they applied for the grant. Evaluators typically contributed to the evaluation portion of the i3 application.

In interviews, the grantees had different opinions about what to look for in an evaluator.¹⁰⁹ “Sometimes with a development grant, you get suckered into working with the expensive nationals,” said one grantee. “We were successful with a small local shop as our evaluator. You don’t need the big company. I have heard of folks spending huge chunks of money on that, but more needs to be spent on the intervention.”

Others disagreed. “I wouldn’t discourage working with a big shop, since they have the ability to turn around documents and find experts in-house,” said another grantee.

“Do they understand What Works Clearinghouse requirements? Have their reports been included in the What Works Clearinghouse?” asked another. “Evaluators need to understand what it means to be in schools,” said another.

“Use references. Have they evaluated a program like yours? Who will do the evaluation? Do they have experience with the mechanics of working with and cleaning the data?” asked one grantee.

Once the evaluator was chosen, several grantees said it was important to maintain contact on an ongoing basis. “Spend a lot of time up front having the researchers learn about the program,” said one. “We had a two-day kick off for the research team to learn about the work. Then we went through the research plan with a fine-toothed comb.”

¹⁰⁵ Institute of Education Sciences, “Low-Cost, Short-Duration Evaluations: Helping States and School Districts Make Evidence-based Decisions,” May 23, 2016. Available at: <https://ies.ed.gov/blogs/research/post/low-cost-short-duration-evaluations-helping-states-and-school-districts-make-evidence-based-decisions>

¹⁰⁶ Institute of Education Sciences, “IES Awards Low-cost, Short-duration Grants to Study Local Programs and Interventions,” July 28, 2016. Available at: https://ies.ed.gov/whatsnew/pressreleases/07_28_2016.asp

¹⁰⁷ This section is based in part on project director answers to the following open-ended interview question: “Were there any particular lessons you learned from how to evaluate a program like yours?”

¹⁰⁸ In this section “success” is defined in two ways: (1) whether an

evaluation was well-conducted, regardless of its findings (positive, null, or negative); and (2) whether the evaluations found positive effects, which is a reflection of the project being evaluated. Both are discussed.

¹⁰⁹ Some related resources include: Institute of Education Sciences, “How To Find A Capable Evaluator To Conduct a Rigorous Evaluation Of An Educational Program Or Practice: A Brief Guide,” June 2007. Available at: <http://coalition4evidence.org/wp-content/uploads/2012/12/PublicationGuideFindingEvaluator07.pdf>; and James Bell Associates, “Evaluation Brief: Locating and Hiring an Evaluator for Your Grant,” July 2007. Available at: <http://www.ibassoc.com/ReportsPublications/Locating%20and%20Hiring%20an%20Evaluator.pdf>

"Have someone on the project team who can interface with the evaluation team and critique and tweak the design to make sure it is a good match for the program work and its assumptions," said another.

In addition to the formal evaluation submitted to the Department of Education, several grantees also conducted separate internal evaluations. "You don't have to completely give up control of evaluating your program to the outside evaluation," said one grantee.

"Our internal evaluator serves as liaison to the external evaluator," said another. "They are gathering a lot of qualitative data and that has been really great, but none of that is in the official evaluation that gets submitted to the What Works Clearinghouse. The Department hasn't figured out a way to capture the full picture while also being rigorous."

The Federal Role

While the independent evaluators played a central role, they did not have complete discretion over the design of their evaluations. The evaluations were subject to review by Abt Associates, a contractor hired to act as a technical assistance provider under the direction of the Institute of Education Sciences (IES).¹¹⁰

As described elsewhere in this report, federal oversight began early in the process. Each i3 project's incoming level of evidence was a deciding factor in whether it received funding, particularly for the larger validation and scale-up grants.

Shortly after receiving the grant, the project evaluators were required to submit a comprehensive evaluation plan that provided details on the evaluation's research questions and methodology. These plans included sections on: evaluator independence, confidentiality protections (including relevant Institutional Review Board (IRB) approvals), descriptions of the evaluated program, chosen research questions, descriptions of comparison

groups and comparison group conditions (no services, business-as-usual services, or an alternative service), a data collection plan, implementation study plans, and analytical methods.¹¹¹

Scale-up and validation grants were expected to meet What Works Clearinghouse (WWC) standards.¹¹² Development grants were expected to "provide evidence of the innovation's promise for improving student outcomes."¹¹³ Ensuring that i3 evaluations met these standards was one of the principal responsibilities of Abt Associates under its contract with IES.¹¹⁴

"Abt was sold to us as a TA provider, but really they were the gateway," said one grantee. "You needed to satisfy them, not just ask them for help. If we didn't do it their way, we had to redo it."

This was an opinion that was widely held among the grantees, but it was not necessarily true. In an interview with Abt representatives and IES, they stressed that their role was advisory. However, their opinions were usually backed by the i3 project officers,¹¹⁵ who wanted the projects to meet high standards and held the grantees accountable for at least achieving the level of rigor specified in their grant applications.¹¹⁶ While some i3 grantees operated under cooperative agreements with the Department of Education, which gave federal officials greater leverage, its authority over local evaluation designs was usually not mandatory, but limited to strong encouragement.¹¹⁷

After the evaluation plans were finalized, Abt maintained contact with the evaluators, tracking evaluation progress, data collection, and implementation through regularly scheduled phone calls.¹¹⁸ In some cases, the project directors sat in on these conversations.

"We participated in all of those calls and it was very important," said one project director. "The external evaluator doesn't always understand our

¹¹⁰ Institute of Education Sciences, "Evaluation of Investing in Innovation (i3)." Available at:

https://ies.ed.gov/ncee/projects/evaluation/assistance_ita.asp

¹¹¹ Abt Associates, "Evaluation Plan Template," August 2016. Available at:

<http://ies.ed.gov/ncee/projects/pdf/EvaluationPlanTemplate.pdf>

¹¹² Institute of Education Sciences, "Procedures and Standards Handbook: Version 3.0," March 2014. Available at:

https://ies.ed.gov/ncee/projects/evaluation/assistance_ita.asp

¹¹³ Government Accountability Office, "Tiered Evidence Grants: Opportunities Exist to Share Lessons from Early Implementation and Inform Future Federal Efforts," September 2016. Available at: <http://www.gao.gov/assets/680/679917.pdf>

¹¹⁴ Institute of Education Sciences, "Evaluation of Investing in Innovation (i3)." Available at:

https://ies.ed.gov/ncee/projects/evaluation/assistance_ita.asp

¹¹⁵ The i3 program was itself being held to high standards by its GPRA standards and pressure from OMB.

¹¹⁶ Some of the grantees missed that standard anyway. Some grantees that had originally planned RCT-based studies in their applications ended up with quasi-experimental designs.

¹¹⁷ Interview with Abt and IES staff, January 4, 2017.

¹¹⁸ According to Abt, project evaluators can expect to engage in an average of 72 calls with a qualified TA provider over the course of their grant. Abt also supports evaluation-related sessions at the annual project director's meeting, runs webinars, and provides tools for evaluations and evaluation plans. Publicly available tools and resources can be found on the IES web site at:

<https://ies.ed.gov/ncee/projects/evaluationTA.asp>

program. Sometimes we interpreted between them. Sometimes we were able to pose questions and there was a shift because of our participation."

According to the Obama administration's FY 2016 budget request, all of the validation and scale-up grants and most of the development grants are on track to meet WWC standards.¹¹⁹ A separate Abt review of the program's first three cohorts also indicated that all of the validation and scale-up grants are using RCTs or QEDs (divided about evenly between the two) and 83 percent of development grants are using RCTs or QEDs (with twice as many QEDs).¹²⁰

As part of its contract with IES, Abt plans to release a final assessment of the i3 evaluations. The first report on the 2010 cohort is expected to be released in early 2017. The review will explore: (1) the extent to which the i3 evaluations are well-designed and well-implemented; and (2) evaluation results for different categories of key i3-funded practices, strategies, and programs.¹²¹

Model Fidelity and Program Implementation

Program implementation was a major focus for every i3 evaluation, particularly the degree to which the tested interventions were faithfully implemented according to their underlying program models. Among other benefits, information on program implementation could shed light on the program's impact findings.

Details on a project's implementation study were included in its Abt-reviewed evaluation plan. Abt also developed a fidelity tracking tool to help grantees to identify their core program components and measures that would determine if they had been implemented consistently.¹²² Nearly all of the final evaluations included sections on implementation and fidelity, some of which went into significant detail.

How well were the projects implemented? Based on interviews and a review of the final evaluations that have been publicly released, most of the i3 grantees fell into one of the following four categories:

- **Well-implemented, Effective Interventions:** The most successful grantees oversaw projects that were both well-implemented and effective, as determined by their impact findings. For these projects, measuring fidelity helped identify core program components that should be implemented faithfully in future program replications.
- **Well-implemented, Ineffective Interventions:** Some grantees ran projects that were well-implemented but produced null results, which suggested shortcomings in the intervention's design. These evaluation results often shed light on options for further improvement.
- **Poorly-implemented, But Possibly Effective Interventions:** For some grantees, the evaluations found poor implementation and little or no impact. In these cases, poor implementation helped explain the poor impact results, leaving open the possibility that the underlying intervention could be effective if implemented faithfully. Projects that fell into this category were often working with low-capacity schools facing resource constraints, insufficient training, high staff turnover, or poor buy-in.
- **Poorly-implemented, Ineffective Interventions:** For a few grantees, poor implementation, poor design, and poor results seemed to go hand in hand. "When something is a problem and isn't working, teachers are going to do what they want to do, which is one of the reasons the fidelity fell off," said one grantee.

When asked, the grantees had varied opinions about the implementation portions of their evaluations, particularly the expectations for model fidelity. "We had a love/hate relationship with fidelity," said one grantee. "That was the hardest part for all of us. If you asked us at the end of year three if we would do things the same way, we would have created an almost entirely different program. But you need to go with the one you invited to the dance."

"There had to be a mind shift change. We are used to continuous improvement and continuous change, but we needed fidelity for three years," said

¹¹⁹ U.S. Department of Education, "Innovation and Improvement: Fiscal Year 2016 Budget Request," pp. G-24-26. Available at: <https://www2.ed.gov/about/overview/budget/budget16/justifications/g-ii.pdf>

¹²⁰ Abt Associates, "Learning from the National Evaluation of i3: Challenges, Responses, and Future Plans," September 11-12, 2014. Available at: http://forumfyi.org/files/Learning_National_Eval_i3.pdf

¹²¹ Institute of Education Sciences, "Evaluation of Investing in Innovation (i3)." Available at: https://ies.ed.gov/ncee/projects/evaluation/assistance_ita.asp

¹²² Government Accountability Office, "Tiered Evidence Grants: Opportunities Exist to Share Lessons from Early Implementation and Inform Future Federal Efforts," September 2016. Available at: <http://www.gao.gov/assets/680/679917.pdf>

one grantee. "That was frustrating."

Others thought the effort was worth it. One used the fidelity measures as a basis for developing a program manual. "This is not a fast food restaurant where everyone knows how to flip a burger," said the project director.

"I understand that a RCT where we got a zero effect is a null finding," said another grantee. "But we did a deep dive on fidelity and fidelity measures and that was the big take away. When we looked at fidelity measures, what we found out was the teachers who implemented with fidelity outperformed. The teachers that didn't did worse than the control group. That is a story I can tell non-researchers."

Fidelity measures could also function as an early warning system for potential problems. The quality of these measures seemed to be an important contributor to a project's overall success. This is discussed further in the Reasons for Success or Failure section of Chapter One.

Tips for Success

What factors were likely to affect evaluation success – either the quality of the evaluation or the likelihood of positive findings? While a full review is beyond the scope of this report,¹²³ the i3 grantees made several suggestions during the interviews.

- **Choose an Experienced Evaluator:** Choosing an experienced evaluator who fully understands and has experience with both What Works Clearinghouse standards and the kind of project being evaluated appears to be critical. Of the 44 final evaluations reviewed for this report, six failed the most basic requirement of generating an equivalent comparison group. Others ran into other problems mentioned below, leaving potentially successful programs with an evaluation finding of no impact. Choosing the right evaluator can, by itself, determine whether a project succeeds or fails.

- **If Possible, Select or Adapt an Intervention with a Strong Research Base:** The level of incoming evidence was a deciding factor for receiving the grant, but several grantees further emphasized its importance as a major driver in their success. This conclusion is reinforced by the higher evaluation success rates of the validation and scale-up grants compared to the development grants, which had lower incoming evidence requirements.
- **Choose an Appropriately Rigorous Evaluation Design:** While evaluations for validation and scale-up grants were expected to meet What Works Clearinghouse standards, several of the development grantees thought the evaluation expectations for their grants were too high.¹²⁴ In particular, several thought RCTs were inappropriate for early-stage innovations. However, some of the development grantees were working with interventions with stronger research bases, which suggests a more varied approach for this category of grants. Projects and evaluators working with genuinely new and innovative projects may want to press Abt and their project officers for more leeway – possibly for the inclusion of a pilot or interim studies before impact studies are conducted in the project's final years.
- **Choose Appropriate Fidelity Measures:** All of the grantees were expected to develop fidelity measures. In interviews, however, some of the development grant recipients said that it was important to spend a few years refining their interventions before finalizing the measures. Fidelity measure quality appeared to be an important contributor to a project's overall success. It is discussed further in Chapter One in the Reasons for Success or Failure section.
- **Choose Appropriate Outcome Measures:** Evaluation results can often be poor if the wrong outcome measures are chosen to assess a program's success. For example, several i3 grantees working on professional development initiatives said that there was too much focus by Abt and federal i3 staff on student

¹²³ There many other resources available on the internet. Some examples include: Institute of Education Sciences, "Designing Strong Studies Webinar Video," July 2014. Available at: <http://ies.ed.gov/ncee/wwc/Multimedia/18>. See also Laura and John Arnold Foundation, "Key Items to Get Right When Conducting Randomized Controlled Trials of Social Programs," February 2016. Available at: <http://www.arnoldfoundation.org/wp-content/uploads/Key->

[Items-to-Get-Right-When-Conducting-Randomized-Controlled-Trials-of-Social-Programs.pdf](#)

¹²⁴ As discussed elsewhere, the decision on evaluation design was negotiated between Abt, the i3 project officers, and the local project, with final decision authority resting with the grantees, although the grantees did not always seem to realize this.

achievement.¹²⁵ "Don't use test scores as the measure unless your program will be pushing that needle fairly closely," said one grantee.

Some grantees addressed this issue with multiple outcome measures. Alternative (or additional) outcome measures chosen in i3 projects included student attendance, GPAs, teacher observations, and teacher knowledge and student attitudes as measured in survey instruments. Some grantees also measured progress through independently administered instruments such as the Peabody Picture Vocabulary Test or Bracken School Readiness Assessment.

Using multiple outcome measures opened up the possibility of mixed results, with successes on some measures but not on others, but it also provided a basis for greater understanding of an intervention's strengths and weaknesses. It also created a layer of added protection in case there were difficulties obtaining needed outcomes data, which happened in some projects (described in the next section).

- **Use Mixed Methods:** In interviews, many of the i3 grantees emphasized the importance of using a mixed methods approach that combined quantitative and qualitative research. "Mixed methods were really useful and focus groups were super important," said one grantee.

"Investing in the qualitative work was really valuable for us," said another grantee. "The evaluation teams would give us briefings 2-3 times per year using quick and dirty updates. That helped identify areas that they thought we should pay attention to."

- **Leave Room for Improvement:** Impact evaluations commonly compare the results of a tested program to business-as-usual practices. Generating positive impacts may be harder, however, if the current practices are reasonably effective and/or the population outcomes are already high.

This was not a problem for most i3 grantees because they were working with disadvantaged populations or in low-performing schools, but this was not always the case. "We are a high

achieving school district already. We have the highest graduation rate in the state," said one grantee. "From that standpoint, we knew we weren't going to see significant growth so we are looking at subgroups."

"We were using self-assessments of teacher efficacy," said another. "Our pre-scores were very high, so there was not a lot of room to grow. After they received the treatment, their self-ratings went down as they learned more about what they didn't know. So our instruments were not strong enough. Our novice principals had rated themselves too high."

- **Choose an Appropriate Comparison Group:** In some cases, the comparison group may not be close enough to the treatment group to be fully comparable. This was a problem for at least two grantees where teaching results for new and inexperienced teachers were being compared to those for far more experienced veteran teachers. Although the two groups achieved comparable results, in evaluation terms this can be seen as a finding of no impact.

- **Be Aware of Potential Control Group Problems:** Assembling an appropriate comparison group was a challenge for many i3 grantees. "There was kicking and screaming in the comparison group," said one grantee. "We would have to turn away children we had served the year before. We would have teachers say we should take kids that needed it."

"We had some who principals who decided to do whatever they wanted," said another grantee. "We would have these randomized groups and then the principals would, on their own, move kids around based on need. Some kids who were not signed up would be put in. One principal told all the control parents they could attend anyway. We found out about it when the numbers didn't match up."

Another potential problem is cross-contamination between the program and comparison group, which can undermine results. "Some of our control kids were – they were all friends – they became mentors to each other. That spilled over into our control group," said one grantee.

¹²⁵ In fact, the choice of outcome measures resulted from a negotiation process. Local projects had substantial authority over this and all aspects of their evaluations. However, both the law creating i3 and the i3 application repeatedly refer to improving academic achievement as a goal of the program.

References to academic achievement are even more explicit in the authorizing legislation for i3's successor, the Education Research and Innovation (EIR) program, which was created by ESSA.

This was not just a problem within schools. It could also occur between schools. "We would sometimes have teachers cross over from a treatment to comparison school. That was an interesting challenge," said another. "The same would happen with students who moved from one school to another."

- **Ensure a Large Enough Sample Size:** In general, larger sample sizes can make it easier to detect smaller program effects. An appropriate sample size can usually be estimated in advance through a statistical power analysis.¹²⁶ Recruiting enough program participants to reach the targeted sample size can often be a difficult challenge, however.

"Building a large RCT sample is harder than you think, even if you think it will be hard," said one grantee. "Budget and be prepared for that. We had to do multiple presentations. Principals, too. It was a big commitment of time just to build the sample. But you need one as big as possible. Don't be on the razor's edge of significance. You don't want to have good results and too small a sample and not be statistically significant."

While larger sample sizes are usually desirable, however, sometimes there can be tradeoffs. "For months we tried to get the evaluators to confront reality. They wanted statistical power. Instead of doing the project, we were dealing with teacher recruitment. We were way too late getting into content," said another grantee.

- **Minimize Sample Attrition:** Attrition is the loss of program participants during a study. Two types of attrition can undermine confidence in a study's results – overall attrition and differential attrition between the program and a comparison group.¹²⁷ Attrition can be a problem, particularly in low-performing schools where there can be high student mobility¹²⁸ and high teacher turnover.¹²⁹

How did the grantees address this problem? "For me, retention was about the importance of relationships," said one grantee. "There's the importance of gentle reminders to the families. Your child will be assessed, so please don't keep them from school. Or reminders to a teacher that we need attendance data for the summer or their curricular diary. We made phone calls, we didn't just send emailed reminders. Some gave us permission to text them."

- **Timing:** Timing issues played a role for several i3 grantees. "While I appreciate the emphasis on the evaluation and results, a three-year timeline feels restrictive," said one grantee. "I am sure we won't see significant changes until year four for evidence of impact. It is the right intervention, but results take longer. It takes a while to stabilize."

"In many cases, this is trailing data," said another. "The results are always two years behind the changes. So we need to look at other data along the way. You need to not just rely on one big outcomes study at the end of the program."¹³⁰

Finally, some programs can show short-term effects, but fade out afterward.¹³¹ By definition, assessing longer-term effects will take more time.

Data

Evaluations, regardless of their design, draw upon a wide variety of data. What types of data did the i3 grantees collect? Where did it come from? What barriers did they face accessing this data, if any?

Their experiences varied. Among the 65 interviewed grantees, slightly more than half (35) reported no major challenges accessing the data that they needed. Just under half (28) reported at least some challenges.¹³²

¹²⁶ Institute of Education Sciences, "Statistical Power Analysis in Education Research," April 2010. Available at: <http://www.ies.ed.gov/ncser/pubs/20103006/>

¹²⁷ U.S. Department of Education, "WWC Standards Brief for Attrition," July 2015. Available at: http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_attrition_080715.pdf

¹²⁸ Sarah Sparks, "Student Mobility: How It Affects Learning," *Education Week*, August 11, 2016. Available at: <https://www.edweek.org/ew/issues/student-mobility/>

¹²⁹ NPR, "Revolving Door Of Teachers Costs Schools Billions Every Year," March 30, 2015. Available at: <http://www.npr.org/sections/ed/2015/03/30/395322012/the-hidden-costs-of-teacher-turnover>

¹³⁰ For suggestions on how to address this issue, see: IES, "Low-Cost, Short-Duration Evaluations: Helping States and School Districts Make Evidence-based Decisions," May 23, 2016. Available at: <http://ies.ed.gov/blogs/research/post/low-cost-short-duration-evaluations-helping-states-and-school-districts-make-evidence-based-decisions>

¹³¹ For an extensive discussion of fade-out effects, see this first in a four-part series: Sarah Sparks, "Focus on Fade-Out: A \$26 Million Project to Hash Out How to Make Pre-K Gains Last," *Education Week*, January 19, 2016. Available at: http://blogs.edweek.org/edweek/inside-school-research/2016/01/fadeout_series_1_education_problems.html

¹³² All of the interviewed grantees were asked the following open-ended question: "Were there any challenges or success stories

For those grantees that experienced challenges, the problems were almost evenly split between local and state education agencies, which were a primary data source. Problems at the local school or district level were usually small and resolvable. However, problems at the state-level were more frequently very significant and they often affected evaluation results when they occurred.

Types of Data Collected

What kinds of data did i3 grantees use? In general, this choice depended on three primary factors: (1) their chosen intervention; (2) its associated evaluation design; and (3) data availability.

Either implicitly or explicitly, most i3 projects and their associated evaluations assumed an underlying logic model that described program inputs, program activities, short-term outputs, and intended program outcomes.¹³³ In most cases, the evaluator used this information to develop a set of evaluation research question(s) and these questions pointed to the kinds of indicators that would be most relevant.¹³⁴ The final decision depended on what data was obtainable, either from existing administrative sources or from new data that was collected as part of the project.

"It's important to know what data is out there in advance," said one grantee. "That can avoid problems. Most districts do not have data on the years of experience of their teachers, for example. The way our city codes things is also different. It helps to know that messy stuff and to have a track record of working with it. If your evaluators don't have that, you are paying them to learn on the job and then it's a budget issue."

Table 3 provides examples of indicators drawn from the 44 final i3 evaluations. They fall into five general categories: (1) school-wide measures; (2) individual student data; (3) personnel-related data; (4) measures of program fidelity and intermediate outputs; and (5) program outcomes. While these indicators varied from project to project, all of the projects included indicators from most or all of these

categories.

In general, the chosen data determined the data source. Many i3 projects relied on administrative data from the schools, particularly school-wide measures, data on school personnel and students, and certain administrative data, such as test results, grades, and graduation rates. Other data came from program implementation, particularly the fidelity measures and some of the chosen outcomes measures, which sometimes relied on surveys or other measures administered as part of the project.

In interviews, some i3 grantees said they felt pressured by Abt and the Department of Education to use school administrative data to measure program outcomes, but many pushed back on this idea. "We think the field would gain from evaluation designs that count for more than just student achievement," said one grantee.

"When you have an intervention that is a few layers removed from the child, the idea that you will impact academics in five years is not realistic," said another.

Those who relied on their own data instruments faced challenges, however. "When we asked about giving our own assessments, that caused panic," said one grantee. "They were already worried about over testing." The project decided not use its own test, possibly to its detriment according to the project director.

"It took time to get those observations at first," said another grantee that was relying on virtual teacher observations with a camera placed in the classroom. "You need to get the teachers familiar with the cameras, turning them on, etc. We made training videos and provided stipends for on-site IT support so it was taken off of the teacher."

Some faced challenges with their measures. "There are not that many tools we could use and few have been normed or validated with this population," said one grantee.

around access and use of data?" Data cited in this section represent the author's interpretation of verbal and written responses. Two grantees did not answer the question.

¹³³ Some good resources on logic models include: W.K. Kellogg Foundation, "Logic Model Development Guide," January 2004. Available at: <http://www.smartgivers.org/uploads/logicmodelguidepdf.pdf>; and Karen Shakman and Sheila M. Rodriguez, "Logic Models for Program Design, Implementation, and Evaluation:

Workshop Toolkit," May 2015. Available at: https://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2015_057.pdf

¹³⁴ Chosen indicators should be well-defined, reliable, valid, measurable, and practical. See: "Selecting Project Indicators," *Monitoring and Evaluation Blog*, May 25, 2013. Available at: <https://evaluateblog.wordpress.com/2013/05/25/selecting-project-indicators/>

Table 3: Examples of Data Collected for i3 Projects

School-wide Data

- Total student enrollment
- Percentage eligible for free/reduced price lunch
- Percentage minority
- Percentages receiving free and reduced price lunch
- Students with disabilities
- School climate measures (surveys, etc.)
- Title I status

Individual Student Data

- Student demographics
 - Race, gender, ethnicity, special education status, age, grade level
- English language learner (ELL) status
- Free/reduced-price lunch status
- Suspensions
- Summer school attendance

School Personnel Data (teachers, principals, school counselors, other staff)

- Certifications, education, and experience
- Qualitative data such as interviews, surveys, focus groups, and observations

Program Fidelity Measures / Intermediate Outputs

- Student participation in tested program (enrollment, attendance)
- Teacher / principal / staff professional development (attendance, time in class, etc.)
- Training observations
- Activity logs, procedural compliance
- Qualitative data such as interviews, focus groups, surveys, and documentation.

Outcomes Data

- Statewide tests (reading, math, science, social studies, etc.)
- Advanced Placement (AP) tests
- ACT scores
- GPA / transcripts
- Graduation / dropout rates
- Proprietary / validated instruments such as:
 - Peabody Picture Vocabulary Test
 - Bracken School Readiness Assessment
- Student surveys
- Teacher surveys
- Teacher evaluations

Source: SIRC review of i3 final evaluations.

Table 4: Data Challenges Experienced by i3 Projects

The following table summarizes data challenges experienced by interviewed i3 grantees as determined by their response to the following open-ended question: “Were there any challenges or success stories around access and use of data?”

Data on the types of challenges experienced exceed the total number of projects facing challenges (28) because some projects cited more than one challenge.

i3 Projects Interviewed	65
<ul style="list-style-type: none"> ▪ No significant data challenges ▪ Some significant data challenges (described below) ▪ No answer 	<p>35 28 2</p>
 Projects Experiencing Significant Data Challenges	 28
<ul style="list-style-type: none"> ▪ Local School or School District Challenges <ul style="list-style-type: none"> • Some local data access difficulties • Consent forms required in some cases • Insufficient local school / district staff capacity • Data cleaning problems • Delays ▪ State Education Agency Challenges <ul style="list-style-type: none"> • Data access difficulties • State test changes • Insufficient state agency staff capacity • Data cleaning problems • Delays ▪ Other Challenges <ul style="list-style-type: none"> • College data access difficulties • Project data vendor problems 	<p>15 3 4 3 2 3 16 7 10 1 1 2 3 2 1</p>

Limitations: Because the data for this table is drawn from an open-ended interview question, it likely under-reports the extent to which these problems were experienced. This data should instead be viewed as the extent to which these were a serious problem. For example, clean data was probably a challenge for more projects than is suggested here, but it was probably not enough of a challenge to be worth mentioning in more than a few cases.

Source: SIRC interview with i3 project directors, summer 2016.

“It is very difficult to assess hands on learning,” said another. “All of the previous tests are multiple choice tests that look at content. We need a portfolio assessment. There are no assessments out there that we can use.”

Qualitative data sometimes helped supplement the quantitative measures. “One of the things we worked on is anecdotal feedback. We did face-to-face interviews because we wanted to know what was really going on,” said another grantee.

Data Access

Obtaining access to needed data was at least somewhat challenging for just under half of the interviewed i3 grantees. The seriousness of these challenges varied. In some cases, they involved temporary delays or could be addressed with workarounds. In other cases, the challenges were more serious, denying access to needed data or altering what was provided enough to potentially affect the evaluations.

Among the 28 interviewed grantees who said they faced challenges, about half (15) reported problems obtaining data from individual schools or local school districts (see Table 4). An almost equal number (16) reported problems obtaining data from state education agencies. Six reported challenges at both levels. Another three experienced problems with obtaining college data or with data vendors.

Of the two major sources of data, state and local education agencies, the state challenges were more serious. Problems encountered with schools or local school districts usually only impacted a small portion of the overall data set. When i3 grantees experienced challenges at the state level, this usually had a much larger impact. The most serious of these challenges involved limited access to state data and changes in state tests. Of the 65 interviewed grantees, about a quarter (16) experienced one or both of these problems at the state level.

At both levels, state and local, major drivers of

data access problems were a concern about protecting data privacy and the associated legal requirements. In the field of education, individual student records are protected by the Family Educational Rights and Privacy Act (FERPA).¹³⁵ In general, FERPA prevents states and local education agencies from disclosing information from a student's education record without the written permission of the student or his or her parent.

However, there are exceptions for research.¹³⁶ Information released for research purposes may include personally identifiable information or alternatively be released in de-identified form (i.e., with identifying information such as the student's name removed),¹³⁷ subject to certain additional requirements such as destroying the data when it is no longer needed for the study.¹³⁸

Subject to these broad protections, the law gives state and local education agencies substantial leeway in deciding on such requests. The extent to which this hindered i3 grantee access to needed data is discussed below.

Data from Local School Districts: At the local level, individual schools were generally willing to share data on participating students. Such cooperation was a condition of the grant and many of the schools already had experience with outside researchers.

- **Local Data Access Difficulties:** Most of the interviewed grantees said they did not experience problems obtaining data from local schools or school districts. Of the three that did, the problems were limited. Two received some, but not all, of the data they requested. The third experienced trouble with control group data, but managed to overcome the problem by obtaining data from the state instead.

The generally high level of success with local schools was due to relationships and project buy-in. “We already had the strong relationships through prior work,” said one grantee, sharing

¹³⁵ 20 U.S.C. § 1232g; 34 CFR Part 99. Information on FERPA can be found at the U.S. Department of Education at <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>. More information is available at the Department's Family Policy Compliance Office (FPCO) at <http://familypolicy.ed.gov/ferpa-school-officials> or its Privacy Technical Assistance Center at <http://ptac.ed.gov/>

¹³⁶ See 34 CFR § 99.31(a)(6) at <https://www.law.cornell.edu/cfr/text/34/99.31>. See also Privacy Technical Assistance Center, “FERPA Exception-Summary,” Available at: http://ptac.ed.gov/sites/default/files/FERPA%20Exceptions_HA

[NDOUT_horizontal_0.pdf](#)

¹³⁷ Privacy Technical Assistance Center, “Data De-identification: An Overview of Basic Terms.” Available at: http://ptac.ed.gov/sites/default/files/data_deidentification_terms.pdf and Reg Leichty and Brenda Leong, “De-Identification & Student Data,” August 2015. Available at: <https://fpf.org/wp-content/uploads/FPF-DeID-FINAL-7242015jp.pdf>

¹³⁸ Privacy Technical Assistance Center, “Best Practices for Data Destruction.” Available at: [http://ptac.ed.gov/sites/default/files/Best%20Practices%20for%20Data%20Destruction%20\(2014-05-06\)%20%5BFinal%5D.pdf](http://ptac.ed.gov/sites/default/files/Best%20Practices%20for%20Data%20Destruction%20(2014-05-06)%20%5BFinal%5D.pdf)

what was a common experience. “We were able to get the data easily.”

“Some of it is they choose you, you can’t choose them,” said another. “The RCT was challenging, but otherwise we set out our expectations with the districts and evaluator and it worked pretty well.”

Those who lacked those relationships, however, could run into problems. “Your project can be really hung up if you don’t have those relationships already in place,” said another grantee. “If you are coming out of the blue, you need to invest the time ahead of time before you write the grant. What data do you want shared and why? Why is it in the interest of the school or others to share it? What assurances are you going to provide?”

“Getting control schools was really tough,” said one grantee that did experience problems. “When we were starting we were going straight to the schools for unidentified data, which was not difficult at first. But as rules became stricter, everyone got nervous. Control schools began to pull out. We had to redo the MOUs so data could go straight from the state to our evaluators.”

Consent Forms: While FERPA allows local jurisdictions to give data to researchers without individual consent, some imposed this requirement on i3 grantees anyway. Four interviewed grantees said they experienced this problem in one or more of the local jurisdictions they were working with (in one case it was for teacher data).

“A couple of schools in one district were concerned and made us get parent signatures, but they were the exception,” said one grantee.

When this happened, however, it could be a problem. “Among our new districts, one required consent. We weren’t able to collect student data because we would need signed consent forms, even with anonymity. We have spent an enormous amount of time on that over 4-6 months.”

Sometimes the challenge was insurmountable. “They required active consent on everything,” said a grantee about one local school district. “We tried to get a high percentage. The most we

got was 38 percent and that wasn’t enough.”

- **Insufficient School Capacity:** For confidentiality and security reasons, most school districts limit access to their data to just a few IT personnel. For this reason, while schools may be willing to provide data, sometimes they do not have the necessary staff capacity.

Three interviewed grantees cited insufficient capacity at the schools as a problem. “We have a mix of school districts that are larger and sophisticated,” said one grantee. Then we have smaller ones where they lack that capacity. It’s not part of their day to day work.”

“We have a wonderful data manager housed inside the schools,” said another that managed to overcome the problem. “He is the one who reaches out to the districts and gets district endorsements and asks for various kinds of data. It was a monster of a task to get the data uniformly across districts that have their own systems and a lot of bureaucracy.”

- **Clean Data:** To be useful, data must be reliable and valid. Two grantees cited this as a problem. “We had to do a lot of work to get clean data. Don’t underestimate the district’s lack of capacity or desire to give you that,” said one grantee.

“The data needed to be scrubbed and cleaned, checked and rechecked. Any evaluator in any district would run into the same issue,” said the other.

- **Delays:** Three grantees said they got the data they needed, but it arrived late. “We did run into challenges, including timing,” said one grantee. “The districts had the info, but they weren’t able to get it to us until after we really needed it to use it.”

“It was difficult to get the accurate data on time,” said another. “A lot of the data comes out in October of the next school year.”

Data from State Education Agencies: Most states operate federally-supported Statewide Longitudinal Data Systems that house data about their state’s K-12 student populations.¹³⁹ Many also link this data to early learning, postsecondary, and workforce

¹³⁹ National Center for Education Statistics, “Statewide Longitudinal Data Systems: Grantee States.” Available at:

<https://nces.ed.gov/programs/slids/stateinfo.asp>

systems.¹⁴⁰

According to a 2014 GAO report, most of these state agencies have established processes that allow researchers who are not employees of the state to propose their own studies for approval.¹⁴¹ Many states also have also adopted research agendas that articulate their research priorities.¹⁴²

Despite this wealth of potential data, however, when i3 projects relied on states they sometimes ran into problems. Of the 28 interviewed grantees that reported data challenges, 16 reported that these included problems at the state level. All 16 reported that these included the most serious challenges: difficulty accessing state data, changes in state tests, or both. The 16 grantees affected by these issues represented about a quarter of the all of i3 grantee interviews (65).

- **State Data Access Difficulties:** Seven of the interviewed grantees said they experienced problems attempting to obtain data from their state education agencies. Several attributed these challenges to an increasingly restrictive political environment. "They don't give anyone student level data anymore," said one grantee about the decisions made in one state. "It's a vast overread of FERPA."

"Our professional evaluators are going to be more careful than any of our school districts. Their livelihood is based on being careful with data," said one grantee. "But there was so much fear."

"We got caught up in the data politics at the state level," said another grantee. "The elected state board has become concerned about the privacy tradeoffs of making data available, arguably to an extreme. They temporarily put a hold on all data requests. Between that and the issue of data quality, our evaluators had to work with us and Abt to figure out alternative data. That was a harrowing moment."

Data access problems may have been worsened by recent concerns about hacking. While the i3

grantees were not asked specifically about data security during the interviews, two raised the issue on their own.

State and local education agencies routinely fend off attempts to hack their data and at least 47 states have laws governing data-breaches.¹⁴³ When interviewed, the two grantees expressed concern that they and other i3 grantees were no less vulnerable to security threats and that the issue needed heightened attention.

- **State Test Changes:** Another common problem for grantees who relied upon statewide test data was changes to the tests. Ten of the interviewed grantees said they experienced this problem.

"The state has changed its high stakes testing," said one grantee. "We could no longer use that as baseline data. The previous test was different from the intermediate test, which is different from the test used now. The data was unusable. The metrics were different."

"One of the challenges for us is that in the middle of the i3 grant, the state changed its standards to Common Core," said another.

"California suspended testing and that caused a lot of problems," said another grantee, citing a challenge that confronted several i3 projects. "We spent enormous hours talking about it. We didn't see any realistic way of getting data for the California schools."

- **State Agency Capacity Limitations, Clean Data, and Delays:** Like local school districts, state education agencies also faced limited staff capacity, problems with clean data, and delays. Three interviewed grantees cited one of more of these problems.

"We had a MOU with the state Department of Education, but when it came to getting the data, it was like pulling teeth," said one grantee.

"We would send someone to physically collect it.

¹⁴⁰ Education Commission of the States, "50-State Comparison: Statewide Longitudinal Data Systems," updated November 16, 2016. Available at: <http://www.ecs.org/state-longitudinal-data-systems/>

¹⁴¹ Governmental Accountability Office, "Education and Workforce Data: Challenges in Matching Student and Worker Information Raise Concerns about Longitudinal Data Systems," November 19, 2004, pp. 21-22. Available at: <http://www.gao.gov/assets/670/667071.pdf>

¹⁴² Ibid. See also: See also: Carla Howe, "State Education

Agencies and Researchers as Partners in Improving Student Outcomes," Brookings Institution, April 20, 2016. Available at: <https://www.brookings.edu/blog/brown-center-chalkboard/2016/04/20/state-education-agencies-and-researchers-as-partners-in-improving-student-outcomes/>

¹⁴³ *Education Week*, "Schools Learn Lessons From Security Breaches," October 19, 2015. Available at: <http://www.edweek.org/ew/articles/2015/10/21/lessons-learned-from-security-breaches.html>

We would get the wrong thing. It was a mess and unusable," she said. "I even asked if we can pay part of a FTE and they wouldn't take the money. A lot of it was fidelity data. That was frustrating. One of the reasons we wanted it was that kids

were moving around the state. We had higher attrition and it would have been less if we had gotten that data."

Chapter Four: Scale

Once a program is proven, how is it sustained or scaled? How are lessons learned more broadly disseminated to the field? This chapter explores each of these issues in depth.

Sustainability

Sustaining any program, including successful ones, can sometimes be a challenge. Like most federal programs, grants made under the i3 program were for a limited duration (five years, although some grantees received no-cost extensions to complete their evaluations).

How are i3 projects being sustained? In general, they have followed one or more of the following strategies:¹⁴⁴

- **Additional i3 Grants:** Several of the grantees applied for additional or subsequent grants from the i3 program. Examples from the 13 grantees with successful final evaluations include: the Children's Literacy Initiative (which jumped from a 2010 validation grant to a 2015 scale-up grant), BARR (which received a development grant in 2010, a validation grant in 2013, and a scale-up grant in 2015),¹⁴⁵ and WestEd (which added a 2012 development grant to its existing 2010 validation grant to test an internet-based version of its Reading Apprenticeship initiative).

Success for All, one of the 2010 scale-up grantees, successfully obtained a 2011 development grant before its 2010 project was complete.¹⁴⁶ However it was excluded in 2012 and 2013 despite being the top-scoring applicant

in the scale-up grant category.¹⁴⁷ After the competitions were over, the Department decided to award no scale-up grants in either year.

Other grantees were unable to obtain additional funds, however, even after work on their original grants was complete. "I was challenged continually by the fact that the priorities kept changing," said one of the fourteen grantees that had achieved positive impact findings in its first grant. "In our naïve way, we really believed that if we proved ourselves, we would get funding. We are in the What Works Clearinghouse. We did apply for a validation award, but the last time school turnaround and early childhood were not there. That's crazy."

- **Other Federal Grants:** While the i3 program is one path to continued federal funding, it is not the only one. Some grantees obtained grants from other programs, such as Promise Neighborhoods or GEAR-UP, although not always for precisely the same work.

In 2013, the Department of Education announced that it had updated its grant requirements to better incorporate evidence. Estimate at the time suggested that the changes would affect over \$2 billion in competitive grants.¹⁴⁸ The Every Student Succeeds Act (ESSA), enacted in late 2015 to replace No Child Left Behind, also established new evidence standards that will also affect many of these grants, including the Education Innovation and Research program that replaced i3.¹⁴⁹ They may also affect state and local funding for school turnaround efforts.¹⁵⁰

¹⁴⁴ This section is based on project director answers to the following open-ended interview question: "What are your plans, if any, for sustaining the initiative once the i3 funding ends?" Additional information was drawn from internal performance reports.

¹⁴⁵ The 2010 grant was received in partnership with the Search Institute and the 2015 grant was with Spurwink Services.

¹⁴⁶ For information about this grant, see: <https://i3community.ed.gov/i3-profiles/32>

¹⁴⁷ Michele McNeil, "Success for All Again Scores Big, And Loses, in i3 Contest," January 17, 2014. Available at: http://blogs.edweek.org/edweek/campaign-k-12/2014/01/success_for_all_wins_then_lose.html

¹⁴⁸ Michele McNeil, "Evidence Matters: U.S. Education

Department Finalizes EDGAR Rules," *Education Week*, August 14, 2013. Available at:

http://blogs.edweek.org/edweek/campaign-k-12/2013/08/evidence_matters_us_education.html

¹⁴⁹ Social Innovation Research Center, "K-12 Education Bill Advances Evidence-based Policy, Replaces i3," December 7, 2015. Available at:

<http://www.socialinnovationcenter.org/?p=1806>; Social Innovation Research Center, "ED Announces First Round of Grants Under i3's Replacement," December 15, 2015. Available at: <http://www.socialinnovationcenter.org/?p=2430>

¹⁵⁰ Alyson Klein, "How Will ESSA Be Different When it Comes to School Turnarounds Than SIG?" *Education Week*, October 25, 2016. Available at: http://blogs.edweek.org/edweek/campaign-k-12/2016/10/how_will_essa_be_different_when_it_comes_to_school_turnarounds_than_sig.html

Grants under the Supporting Effective Educator Development (SEED) grant program must be evidence-based.¹⁵¹ The Department issued initial non-regulatory guidance on ESSA's evidence provisions in September of 2016.¹⁵²

- **Other State and Local Funding:** A few i3 grantees said that their i3 work made it easier to apply for state and local grants because they had better research evidence that could be included in their grant applications.
- **Continued School Funding / Fee-for-Service:** In several cases, the grantees planned to support their ongoing work on their own, either because the grantees were schools or because they charged fees to their partner schools.

"We want the district to sustain it at their own expense," said one grantee. "The district has embraced it by purchasing the training and materials for other grade levels. Teachers outside the projects have been trained in it. The school district itself has been rewarded with other grants. This was a school district that had one of the lowest graduation rates in the county. As a result of the changes, they are no longer at the bottom. Their teachers apply for awards. There are benefits for them and lots of spin-offs."

"We grew the number of schools beyond the i3 schools," said another grantee with a successful evaluation results. "At the district level, they wanted to go to other schools and they were willing to spend their own resources for that."

- **Sustained Philanthropic Support:** Some i3 grantees may be able to achieve sustained support from the foundations and other philanthropic individuals and institutions that helped them meet their i3 match requirements.

It is not clear how successful they will be, however, since such support (at least from foundations) has generally been difficult for nonprofit organizations to sustain.¹⁵³ Comparable philanthropic support for grantees under the Social Innovation Fund, which

launched at the same as i3, usually dropped over time.¹⁵⁴

This appears to be true for i3 as well. According to grantees who were familiar with the i3 Foundation Registry,¹⁵⁵ there appears to be less interest in supporting i3 grantees now compared to the program's early days. "In the first round there were a bunch of funders. By 2-3 years later, there were many fewer," said one grantee that did fundraising for two separate i3 grants.

"It puts a lot of pressure on a nonprofit," said another grantee. "Many of our cohort members were school districts. We are a small nonprofit. There is intense pressure on us to sustain the funding that government grants allow us."

- **Projects Continued without Substantial i3 Grantee Involvement:** In some cases, i3 grantees planned to let local schools continue on their own, with little ongoing support. "We hope the activities put in place will continue rolling. There seems to be enough excitement about those things that they will continue," said one grantee. "I hope and fully expect that the processes and partnerships will become ingrained in the schools and when the project is over they will keep doing it," said another.

In some cases, the grantees intended to stay in touch, but provide only modest support. "Most of them we will continue to work with, but at a low level which is our preference," said one. "Be cautious about making the grant personnel heavy. That is not easy to sustain post-grant. You need to go into your grant thinking about what can be sustained long term," she said.

- **Projects That Are Not Sustained:** Finally, some i3-funded projects either have not been sustained, may not be, or if so are likely to undergo major changes. These were typically projects that did not achieve significant positive results in their independent evaluations.

[k-12/2016/10/essa_different_SIG_school_turnarounds.html](https://www2.ed.gov/policy/elsec/leg/essa/guidanceeusesevinvestment.pdf);

¹⁵¹ Information on the SEED grant program is available at: <https://www2.ed.gov/policy/elsec/leg/essa/guidanceeusesevinvestment.pdf>

¹⁵² U.S. Department of Education, "Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments," September 16, 2016. Available at: <https://www2.ed.gov/policy/elsec/leg/essa/guidanceeusesevinvestment.pdf>

¹⁵³ Center for Effective Philanthropy, "Nonprofit Challenges: What

Foundations Can Do," May 2013. Available at:

<http://www.effectivephilanthropy.org>

¹⁵⁴ Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, pp. 30-31. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf

¹⁵⁵ For more information on the i3 Foundation Registry, see its web site at: <https://www.foundationregistryi3.org/>

Dissemination

Dissemination of program results has been a consistent feature of the i3 program. Plans for dissemination were among the criteria included in the grant applications and grantee progress has been tracked in grantee reports to i3.

How well has this dissemination worked? Common activities included the following:

- news coverage in local media and *Education Week*, a national education trade publication;
- journal articles;
- presentations at national and state meetings and conferences;
- organizational newsletters and articles and blogs on the organizational web site;
- articles posted on the Department of Education's Office of Innovation and Improvement¹⁵⁶ and i3 Learning Community¹⁵⁷ web sites;
- submission of the final evaluation to ERIC, the Department of Education's online library of education resources;¹⁵⁸ and
- incorporation of evaluation results into grant applications.

In interviews, however, it appeared that the level of energy put into these efforts depended heavily upon the success of the projects.¹⁵⁹ "We did the minimum that we were required to do to because our evaluation didn't show impact," said one grantee.

Others did more. "We have gotten a couple of articles in professional journals. I have a blog. We have a web site and use Twitter and Facebook intermittently. We could do better," said one grantee. "My real struggle is understanding why we are doing this. Clearly ED wants it. A funder is happy when you publish or present, but beyond that it gets a little fuzzy. What are we hoping to accomplish?"

"We are doing our best. We have an outreach Department that does marketing and roams around doing presentations and things you would do. We're better than most nonprofits," said another.

Some were more active, distributing their results through organizational networks or communicating directly with state and local education agencies and

policymakers, but others thought the Department of Education could take a more active role.

- **Department of Education's Role:** While many of the grantees took advantage of the dissemination opportunities provided on Department of Education web sites (listed above), some thought that i3 and the What Works Clearinghouse could be doing more.

"I would really like to see i3 take a more active role in the dissemination of grantee work," said one. "It would behoove i3 to look at their own strategy. Maybe do a meta-analysis approach," said another. "They need to disseminate to decision makers. As it is, each project does it on their own. So a lot of valuable lessons learned don't get communicated up the chain."

Some of the i3 evaluations have been included in the What Works Clearinghouse (see Appendix B) and this may become increasingly important as new evidence requirements are incorporated into Department of Education discretionary grant programs.¹⁶⁰

Some grantees expressed frustration with getting their results reviewed by the What Works Clearinghouse, however. "There is a years-long waiting list," said one. "We heard two years ago that they have a lot of studies to look at. What's the plan there?"¹⁶¹

- **Organizational Networks:** Many of the nonprofit grantees were part of larger networks, which gave them an advantage in dissemination. "Because of the national network, we are able to use the information in other states and with Congress. Now we are in a place where the states are coming back to us. We have six years under our belt with lots of great examples," said one grantee.

"When the report came out there was a flurry of emails from all over the world," said another with international connections. "They are looking for lessons learned from our involvement with i3," said another member of the same team.

¹⁵⁶ See <https://sites.ed.gov/oii/>

¹⁵⁷ See <https://i3community.ed.gov>

¹⁵⁸ See <https://eric.ed.gov/>

¹⁵⁹ This section is partly based on project director answers to the following open-ended interview question: "What are your plans for disseminating information about your grant, if any?" Internal performance reports were also a significant source of information.

¹⁶⁰ See also Michele McNeil, "Evidence Matters: U.S. Education Department Finalizes EDGAR Rules," *Education Week*, August 14, 2013. Available at:

http://blogs.edweek.org/edweek/campaign-k-12/2013/08/evidence_matters_us_education_.html

¹⁶¹ The What Works Clearinghouse does not review studies upon request. Its reviews are based on pre-existing protocols that do not prioritize i3 or EIR projects.

The largest grantees also had substantial dissemination reach within their own organizations. "We have 1,000 full time employees," said another grantee. "So we need to double down on that internally."

- **Communication with State and Local Policymakers:** Some of the grantees communicated their results directly to partners and officials with authority over policy or grantmaking. "We have strong support from policymakers in our states," said one grantee.

Others faced restrictions on such outreach or other barriers. "Our biggest challenge is disseminating among district leaders and state leaders. Those are our target audiences, but there is so much turnover that it is hard to sustain communications," said another.

Scale

Disseminating evidence is an important first step, but convincing state and local education agencies to use such evidence is a more challenging task. The i3 program was designed to do both.

How well have i3's scaling efforts worked? What lessons have been learned?

Progress of Scale-up Grants

Four scale-up grants ranging from \$45 to 50 million were made in the program's first year and their results are now in. Among this group, all four expanded their evidence-based initiatives, although some did not reach their expansion targets.¹⁶² Two did so with positive impact results while the other two did so with mixed results.

- **KIPP:** The KIPP charter school network expanded by 48 schools, from 82 in 2010 to 130

in 2014. Its RCT-based evaluation found that KIPP elementary schools have a positive impact on student reading and math achievement and that KIPP middle schools have positive effects on math, reading, science, and social studies.¹⁶³

- **Reading Recovery / Ohio State University:** This program provided one-on-one Reading Recovery lessons to 61,992 students in over 1,300 schools. Its RCT-based evaluation indicated that it was well-implemented and it found positive program effects on student reading comprehension.¹⁶⁴
- **Success for All:** Through the fourth year of this grant, SFA's whole school turnaround program was launched in 447 new schools and reached an estimated 276,000 students. Its RCT-based evaluation found mixed results, including positive effects on student phonics and pre-literacy skills, but no effects on reading comprehension, special education designations, or rates at which students were held back to repeat a grade.¹⁶⁵
- **Teach for America:** From 2010 to 2015, the number of Teach for America (TFA) corps members grew from 7,352 to 10,500.¹⁶⁶ However, its RCT-based evaluation found mixed results, with positive effects on reading for students in grades K-2 and math in grades 1 and 2, but otherwise no difference between TFA teachers with an average of 1.7 years' experience and veteran teachers with an average of 13.6 years' experience.

All four scale-up grantees grew during their grant periods. It is less clear how much they would have grown without their grants, information that would cast additional light on the importance of i3.

All four had been growing prior to receiving i3 funding, so it is possible that they would have grown anyway. Additional insights could be gained from a

¹⁶² In interviews, some attributed this to the unanticipated effects of the 2008-2009 recession and its impact on state and local education budgets. The Department of Education also noted this in its FY 2007 budget request. See U.S. Department of Education, "Innovation and Improvement: Fiscal Year 2017 Budget Request," p. F-35. Available at: <https://www2.ed.gov/about/overview/budget/budget17/justifications/f-ii.pdf>

¹⁶³ Christina Clark Tuttle, et al, "Understanding the Effect of KIPP as it Scales: Volume I, Impacts on Achievement and Other Outcomes," Mathematica Policy Research, pp. A3-A8, p. 31. Available at: <http://www.kipp.org/results/independent-reports/#mathematica-2015-report>

¹⁶⁴ Henry May, Philip Sirinides, Abigail Gray, and Heather Goldsworthy, "Reading Recovery: An Evaluation of the Four-

Year i3 Scale-Up," Consortium for Policy Research in Education, March 2016, pp. 2, 11. Available at: <http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>

¹⁶⁵ Janet Quint, et al., "Scaling Up the Success for All Model of School Reform," MDRC, September 2015, p. iii. Available at: <http://www.mdrc.org/publication/scaling-success-all-model-school-reform>

¹⁶⁶ Sara Mead, Carolyn Chuong, and Caroline Goodson, "Exponential Growth, Unexpected Challenges: How Teach for America Grew in Scale and Impact" Bellwether Partners, February 2, 2015, p. 74. Available at: <http://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact>

review of the growth patterns for applicants that came close, but did not receive i3 scale-up grants.

Absent such information, the best that can be said is that all four did grow and the addition of \$45-50 million in federal grants probably made a large difference.

These grants and their associated evaluations also generated several important lessons, including: (a) the importance of strong intermediary organizations when growing evidence-based programs; and (b) the unique challenges faced when taking programs to scale.

The Need for Strong Intermediaries

Prior research has suggested that effectively scaling evidence-based programs in schools may require the active involvement of strong intermediaries.¹⁶⁷ This role can be played by leaders within the schools, external organizations with connections to the schools, or both.

Education practitioners often show an interest in research,¹⁶⁸ but they also are often overwhelmed and do not have the time or expertise needed to keep up with the latest developments in the field.¹⁶⁹ Instead, they rely on information gathered from social networks, including peers and trusted external organizations such as professional associations.¹⁷⁰ Absent the leadership of such knowledge brokers, the diffusion of evidence-based programs in schools is typically slow or non-existent, particularly in low-performing schools.¹⁷¹ When evidence is used, it is usually to justify the continuation of existing policies or practices.¹⁷²

The literature suggests that intermediaries may play a crucial role in overcoming these barriers. The i3 program represents, in part, an experiment in the use of such intermediaries to help spread the use of evidence-based programs and practices. It has tested this proposition by providing grants to competitively chosen external organizations or local

school districts, both of which played intermediary roles with their partner schools.

How well has this strategy worked? Preliminary evidence from the i3 program seems to confirm the importance of such intermediaries. There are at least two indicators of this.

- **High-capacity Scale-up Grantees Were More Successful than Local School Districts:** One indicator of the importance of intermediaries in i3 is the success that scale-up grantees experienced compared to development grantees, particularly the local school districts.¹⁷³

While all four of the scale-up grantees achieved either mixed or positive impact in their schools, this was true for only about a quarter of the development grantees (8 of 30) and local school districts (4 of 16).

Development grantees had lower levels of incoming evidence, which partly explained the difference. However, several were also clearly hindered by basic capacities in program design, execution, and evaluation.¹⁷⁴ These disparities were most evident among the local school districts.

Moreover, the comparably worse performance among local school district grantees came despite winning highly competitive i3 grants, which presumably made them better positioned to succeed than the schools that the scale-up grantees were working with.¹⁷⁵

- **Scale-up Grantees Had a National Focus and Broader Reach:** The importance of intermediaries to scaling evidence-based programs could also be seen in the kinds of work being done by intermediaries compared to local school districts.

The scale-up grantees worked with schools spread across the country and made substantial

¹⁶⁷ Additional resources on the dissemination of evidence-based practices can be found at the National Center for Research in Policy and Practice at <http://www.ncrpp.org/resources>

¹⁶⁸ National Center for Research in Policy and Practice, "Research Use in Large School Districts," March 1, 2016. Available at: http://ncrpp.org/assets/documents/NCRPP_Research-Use-in-Largest-Districts.pdf

¹⁶⁹ Kara Finnigan and Alan Daly, *Using Evidence in Education: From the Schoolhouse Door to Capitol Hill* (New York: Springer International Publishing, 2014), p. 178.

¹⁷⁰ *Ibid.*, pp. 3, 14, 87, 101, 109-113, 181.

¹⁷¹ *Ibid.*, p. vii, 21, 27-28.

¹⁷² *Ibid.*, p. 35.

¹⁷³ All but one of the local school districts were development grantees.

¹⁷⁴ This was evident from internal performance reports, interviews, and implementation studies, the last of which were public information and part of the final evaluations. More on this topic can be found in the Capacity Building section of Chapter One.

¹⁷⁵ The Department received almost 5,000 i3 applications or pre-applications between 2010 and 2015, but made only 156 grants, for a total acceptance rate of 3.1 percent. See U.S. Department of Education, "Innovation and Improvement: Fiscal Year 2017 Budget Request," p. F-31. Available <https://www2.ed.gov/about/overview/budget/budget17/justifications/f-ii.pdf>

investments in outreach and support for their national networks (discussed more below).

By contrast, the local school districts only worked with schools in their own districts. Even among the three local school districts with positive evaluation results, they only seemed to engage in traditional dissemination efforts as required by their grants. None seemed likely to engage with other schools more proactively, at least not at local taxpayer expense.

If local school personnel were to engage with other schools, this would likely be through a newly-created nonprofit intermediary or external partner – as occurred, for instance, with the BARR program, a successful development grant.

The Challenges of Taking Programs to Scale

Another broad lesson could be found in the significant challenges posed by scaling an evidence-based program. While these challenges were similar to those faced by other i3 grantees – program launch, school partnerships, implementing programs with fidelity, and capacity building – they tended to become more difficult and change as a program was more widely implemented.

For example, as a program is more widely adopted, it is no longer operated or overseen by its original designers, the people who are most familiar with its intricacies. It may also need to be adapted to differing local contexts and different populations. Program operators may not always have the luxury of picking and choosing among high-capacity partners. Relationships, while important, can no longer be assumed and often do not run as deep.

How did the i3 scale-up grantees handle these challenges? Some of the lessons drawn from their evaluations were these:

- **Program Marketing:** The first step in scaling a program is broader market awareness. This was less of an issue for lower-tier development grants

because the participating organizations, schools, and project personnel often already knew one another.

When projects go to scale, however, this can take more work. Some of it includes the passive dissemination activities described earlier in this chapter, such as conferences and journal publications, but most of it is more proactive, requiring face-to-face meetings with potential partners.

“It’s slow going because it’s so personal,” one director told the evaluators for the Reading Recovery grant, “so I spend a lot of time on the road and in hotels.”¹⁷⁶

“You have to make a personal connection,” said another director. “Somebody’s not going to wake up some morning and decide ‘This is what I’m going to do. I’m going to do Reading Recovery.’ It’s that personal connection that you make to recruit people.”¹⁷⁷

Success for All leaders believed that word of mouth was the most effective way to engage new schools.¹⁷⁸ The prestige of the i3 grant may have also helped.¹⁷⁹

- **Closing the Deal:** Making contact with a school or other potential partner was usually just the first step. The final decision was usually a simple cost-benefit judgement based on the services being offered, the need for those services, the capacity of the school or partner to handle the initiative, and any associated costs.

Success for All’s leadership believed that a proposed project’s distinctiveness also helped.¹⁸⁰ If a new program was not sufficiently different from what schools were already doing, they might not see its value. The internal capacity of the schools, which often experienced high staff turnover and sometimes faced the threat of closure, was also an issue.

¹⁷⁶ Henry May, Philip Sirinides, Abigail Gray, and Heather Goldsworthy, "Reading Recovery: An Evaluation of the Four-Year i3 Scale-Up," Consortium for Policy Research in Education, March 2016, p. 12. Available at: <http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>

¹⁷⁷ Ibid.

¹⁷⁸ Janet Quint, et al., "Scaling Up the Success for All Model of School Reform," MDRC, September 2015, p. 139. Available at: <http://www.mdrc.org/publication/scaling-success-all-model-school-reform>

¹⁷⁹ Henry May, Philip Sirinides, Abigail Gray, and Heather Goldsworthy, "Reading Recovery: An Evaluation of the Four-Year i3 Scale-Up," Consortium for Policy Research in Education, March 2016, p. 19. Available at: <http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>

¹⁸⁰ Janet Quint, et al., "Scaling Up the Success for All Model of School Reform," MDRC, September 2015, p. 139. Available at: <http://www.mdrc.org/publication/scaling-success-all-model-school-reform>

For many, cost was the final deciding factor, particularly in the aftermath of the 2008-2009 recession when state and local budgets were tight. When services were not being offered for free, school administrators could find it difficult to fit new services into their tight budgets.

Success for All subsidized most of its school partners, for example, but several still decided not to participate in its initiative because of the cost.¹⁸¹

This was also a factor for a validation grantee that has not yet released its final evaluation. "All of these reforms can be done with Title I money, but it is seen as the most precious discretionary money they have," said the project director. "On-the-ground budget reallocations are a tough fight. They'll ask 'so you want me to let a counselor go?' It's a Hobson's choice sometimes. They are underfunded to begin with."

- **Obtaining Broad Buy-in:** While initial sign-off on a project usually comes at the district or school principal level, effective implementation often requires broader buy-in by teachers and other school staff. This was not always easy.

"The need to recruit so many schools in such a short timeframe made it difficult to establish the school-level buy-in and relationships for strong initial implementation," wrote one grantee in its performance report to the Department of Education.

Buy-in affected not just implementation generally, but also fidelity to the tested intervention, which could affect both program outcomes and evaluation results.

"Many teachers did not appreciate beforehand the extent to which SFA is scripted," wrote SFA's evaluator in one report. "Teachers were inclined to complain that it stifled their creativity and, on the teacher survey, were likely to agree that their reading program was too rigid or too scripted."¹⁸² To address this, Success for All requires 80

percent of a school's teachers to vote to adopt it before it will proceed with a project.

Buy-in was also important to the other scale-up grants. Staff for the Reading Recovery scale-up found ways to test its partners' commitment.

"One of the things we also put in place around i3 is that we ask our schools to fill out a five or six-page application," said one director. "It's just to ensure that that the school doesn't just say, 'Oh yeah! I'm going to train a teacher. Give me this money.' [We want to ensure that] they put some thought into it.

"It's a positive thing but at the same time it could be impacting the number of schools who are making investments. We had one school say, 'I'm not filling that thing out.' So, we didn't get them."¹⁸³

- **Capacity Building:** Some of the i3 grantees worried that scaling would make each additional program or recruit harder to work with while maintaining high standards.

"When you're small, you're going after the lowest-hanging fruit," said one executive for Teach for America. "We're probably already getting the easiest 5,000 applicants. Each increment after that will be harder to get and more labor-intensive."¹⁸⁴

Teach for America addressed this problem by building its organizational capacity. Among other investments, it expanded its use of recruiting teams and created an electronic data tracking system. This technology investment allowed the organization to analyze impact data on recruits to determine which competencies best predicted later classroom success.

"Scale and quality aren't necessarily countervailing forces—but maintaining quality while growing requires intentional focus matched by resources," wrote the authors of a report on TFA's scaling efforts.¹⁸⁵

¹⁸¹ Ibid., p. 124.

¹⁸² Janet C. Quint, et al., "The Success for All Model of School Reform: Early Findings from the Investing in Innovation (i3) Scale-up," MDRC, October 2013, p. 38. Available at: <http://www.mdrc.org/publication/success-all-model-school-reform>

¹⁸³ Henry May, Philip Sirinides, Abigail Gray, and Heather Goldsworthy, "Reading Recovery: An Evaluation of the Four-Year i3 Scale-Up," Consortium for Policy Research in Education, March 2016, p. 23. Available at:

<http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>

¹⁸⁴ Sara Mead, Carolyn Chuong, and Caroline Goodson, "Exponential Growth, Unexpected Challenges: How Teach for America Grew in Scale and Impact" Bellwether Partners, February 2, 2015, p. 31. Available at: <http://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact>

¹⁸⁵ Ibid., p. 78.

Building capacity costs money. The scale-up and validation grants were sizable and allowed significant investments in capacity, but they are not enough by themselves. These i3 grantees faced many of the same sustainability concerns

faced by other i3 grantees and they often used the same strategies, although some may have made more sizable investments in their ability to influence public policy.¹⁸⁶

¹⁸⁶ Ibid., p. 56.

Epilogue

Viewed in isolation, the i3 program – while imperfect – appears to have been generally successful. But the newly-renamed Education Innovation and Research (EIR) program does not exist in a silo.

The program exists in a larger policy context – one that includes other federal programs, state and local education agencies, and perhaps most importantly, a political context that has become very different in the aftermath of the 2016 elections.

How does EIR fit into this larger context? As one administration exits the stage, what fate awaits it in the next one? If the new Republican administration and GOP Congress choose to keep it, how might it change?

New Administration, New Priorities

In late 2015, Congress reworked i3 as part of the Every Student Succeeds Act, bipartisan legislation that replaced No Child Left Behind. The new EIR program now bears a bipartisan imprint.

Still, as a program originally associated with the Obama administration, its fate under the incoming administration is uncertain. The Trump administration may decide to eliminate it, but it may also see it as useful tool for furthering its school choice and accountability agendas. Moreover, support for evidence-based policy and tiered evidence initiatives (like EIR) more generally has been growing among Republicans on Capitol Hill.

When asked, Rick Hess, the Director of Education Policy Studies at the American Enterprise Institute, began by making the pessimistic case. “I don’t know what will be done in the new administration,” he said, “but I assume the new administration and Congress will take a hard look at the full array of Obama initiatives, including this one.”¹⁸⁷

“If i3 had happened outside the context of Race to the Top and had not been locked arm-in-arm with foundations on Common Core, I think I would have looked upon it differently, because historically the idea of public-private partnerships has a lot of appeal.”

“On the other hand, there are folks who are interested in school choice. They might see it as a vehicle for encouraging more choice programs and expanding efforts to study their impact,” he said.

The president-elect has pledged \$20 billion for school choice. His Education Secretary-designee, Betsy DeVos, is a strong school choice and accountability advocate.¹⁸⁸ Federal support for charter schools is overseen by the Office of Innovation and Improvement, the same division that runs EIR.¹⁸⁹ As a program that provided support for charter school initiatives like KIPP and New Schools for New Orleans, EIR could be useful to the incoming administration.

Others point to potential support in Congress. “Betsy DeVos will be extremely important, but there is also more appetite on the congressional side than there used to be,” said Grover “Russ” Whitehurst, a former director of the Institute of Education Sciences under President George W. Bush.¹⁹⁰

“There is still bipartisan momentum to increase evidence use within federal policy,” agreed Martin West, associate professor at the Harvard Graduate School of Education and a former senior education policy advisor to Sen. Lamar Alexander (R-TN). “I think the status of programs like EIR hinges on that broader momentum.”¹⁹¹

Support for evidence-based policy has been growing among Republicans in recent years. In the summer of 2016, House Speaker Paul Ryan and other members of the House Republican leadership made evidence a central component of a policy plan

¹⁸⁷ Interview, January 13, 2017.

¹⁸⁸ Michael McShane, “Betsy DeVos, the (Relatively Mainstream) Reformer,” *Education Next*, Vol. 17, No. 3. Available at: <http://educationnext.org/betsy-devos-relatively-mainstream-reformer-education-secretary/>

¹⁸⁹ Office of Innovation and Improvement, “Charter Schools Program.” See

<https://www2.ed.gov/about/offices/list/oii/csp/index.html>

¹⁹⁰ Interview, January 9, 2017.

¹⁹¹ Interview, January 12, 2017.

that it ran on in the fall, called A Better Way.¹⁹² Although the plan did not mention EIR specifically, it endorsed tiered-evidence initiatives in general.

The broader focus on evidence has also drawn cautious support from analysts at conservative organizations like the American Enterprise Institute, Manhattan Institute, and Heritage Foundation.¹⁹³ In November, the Heritage Foundation endorsed the increased use of evidence in the federal budget process.¹⁹⁴

Given these varied sources of potential support, EIR's fate is unclear. If it were to be kept in place, however, how might it be changed?

Where Does EIR Belong?

One issue that has emerged is the program's organizational placement. EIR is partly an education research program. Does it belong in the Office of Innovation and Improvement (OII), where it is located now, or at the Institute of Education Sciences (IES)?

"I think it belongs on the program side at the Department," said Nadya Dabby, the last Obama appointee to oversee OII.¹⁹⁵ "Many of i3's greatest contributions extend beyond using or generating evidence."

"The thing I heard over and over from grantees is that it changed how their organizations think about data, improvement and evidence—beyond their i3-funded project. You only get that kind of organizational change if the work at its core is led by practitioners and not the researchers," she said.

"IES did the evidence reviews, but we put the mechanics in OII. That made sense to us," said Robert Gordon, who previously served at both the Department of Education and OMB during the Obama administration and helped design the program. "We still got tons of expertise from IES."

What were the benefits of putting the program in OII? "It is no knock on IES – what they are doing is incredible and really pushing the evidence movement forward," said Shane Mulhern, the program's most recent director. "However, EIR is about evidence in

practice. We want those grantees to be in a continual cycle of learning and improvement."

"It shouldn't matter where the evidence comes from, but that's not the way human behavior works – people leave evidence on the shelf unless they have some personal connection to it," said Dabby. "i3 has changed practices and efforts in schools and education nonprofits across the country. You only get that if the primary 'clients' of the program are schools and nonprofits."

Others think the program should be moved. "I was the director of IES. You need a wall between you and politics," said Whitehurst.

"EIR ought to be run by IES. OII could easily be politicized and I don't think anyone wants these funds to be compromised by the sense that someone has their thumb on the scale," he said. "That has not been true so far, but it would be better in IES."¹⁹⁶

"Perception of politicization also matters," said West, the former advisor to Sen. Lamar Alexander. "Within the Department there are strong pressures to align everything with the overall policy priorities. This is supposed to be bottom-up and field generated. What you want is a more open-ended process, with no absolute priorities."

"I would argue that i3 and EIR are research programs and belong in IES because it is non-partisan and non-political," said Ruth Neild, the most recent director of IES. "The lesson of i3 is that there is a hunger for evaluation among organizations that do not have the research expertise to compete successfully for IES research grants. They need an easier on-ramp – but it should be housed in a scientific agency with protections for independence and a staff with research training," she said.¹⁹⁷

"I don't have strong opinions about this as long as attention is paid to lessons we learned about what works well and is successful," said Jim Shelton, a former Deputy Secretary of Education who also oversaw the program in its early years.¹⁹⁸

"If they can build that at IES that's good. If it's at OII, that's fine. I am less concerned about where and

¹⁹² Patrick Lester, "Republicans Deploy an Old Tool in Combating Poverty: Evidence," *Government Executive*, June 14, 2016. Available at: <http://www.govexec.com/excellence/promising-practices/2016/06/republicans-see-silver-bullet-combating-poverty-evidence/129068/>

¹⁹³ Ibid.

¹⁹⁴ Patrick Lester, "Heritage Foundation Endorses Increased Use of Evidence in Budgeting for Trump Administration," December 7, 2016. Available at:

<http://www.socialinnovationcenter.org/?p=2344>

¹⁹⁵ Interview, January 12, 2017.

¹⁹⁶ For a discussion of the statutory protections for independence provided to IES see: Grover J. Whitehurst, "Rigor and Relevance Redux," Institute of Education Sciences, November 2008, pp. 5-6. Available at:

<https://ies.ed.gov/director/pdf/20096010.pdf>

¹⁹⁷ Interview, December 28, 2016.

¹⁹⁸ Interview, January 13, 2017.

more concerned about capacity," he said.

Innovation

As discussed elsewhere in this report, one of i3's main goals was to stimulate and leverage innovation in education. So far, however, it appears to have been less effective at this than its other goals. How might its work on innovation be improved?

"I'm not shocked that validation and scale-up grants were more successful. We were criticized for the same old, same old," said Gordon. "My reaction always was that this is a government program. Identifying initiatives with strong evidence and building them is a relatively straightforward task for government to perform well."

"Identifying early stage brilliant ideas is less objective and requires a different kind of insight. That's not to say government can't do it well, but it is more difficult," he said.

Others were more critical. "Yes, research is a business you want government to be in," said Hess of the American Enterprise Institute. "It is a public good and it is an appropriate role for the national government."

"But there is a difference between that and development. In medical research, NIH spends \$40 billion on bench science. That's people in labs figuring out the building blocks that later gets monetized by other actors who turn it into drugs. We don't want government doing the second part," he said.

"The danger is that i3 was intended to leverage philanthropy. You wind up having the feds shoulder-to-shoulder with specific philanthropic agendas," he said. "It is easy to talk about programs that were supported. What isn't noticed was what did not happen or was discouraged."

"Research is an appropriate federal role, it is a public good," said Whitehurst, the former IES director under President Bush. "But the priorities should not be exclusively or primarily the administration's priorities, they should come from the research and practice communities along with political players among which a presidential administration is but

one."

"The people closer to the work have a lot more knowledge than career bureaucrats," he said. "A state department of education is not going to turn around schools, and federal officials aren't either. If anything is top-down, it should come from legislation, not bureaucrats."

"I share Russ's concern about innovation being a poor competency for the feds," said West. "Deciding which innovations are worthy of encouragement is not a natural role for them. But I am reluctant to recommend that it should be eliminated altogether. The development-validation-scale process should be more closely linked. There should be more of a pipeline approach."

Obama officials overseeing the program were keenly aware of how difficult spurring innovation can be. "One of the things I struggled with was when I was looking for expert reviewers," said Shelton. "I leaned toward putting them on scale-ups. In hindsight, I should have put them on the development grants. There was already a significant track record on the scale-ups. Recognizing something that doesn't have a track record is harder."

"We were forced to make decisions based on reading reports. It is difficult to find an early stage investment firm that does not deeply engage with an organization they are planning to invest in," he said. "In some cases, we need to test using intermediaries. They are in the business of finding and supporting those organizations."¹⁹⁹

One other idea proposed by the Obama administration, but never approved by the Republican-controlled Congress, was a new initiative within i3 modeled after the Defense Advanced Research Projects Agency (DARPA). Called ARPA-ED, the program would have focused on a few high-value research projects using technology.²⁰⁰

The proposal drew early qualified support from Hess.²⁰¹ It may draw greater support from the GOP Congress if a variant is proposed by the incoming administration. "On the researcher side, we need a DARPA that moves quickly and allows risky investment, not five year grants," said Whitehurst. "We need that on the researcher side of things."

¹⁹⁹ This was the approach used by the Social Innovation Fund. See Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, p.p. 19, 34-38. Available at: http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf

²⁰⁰ U.S. Department of Education, "Winning the Education Future:

The Role of ARPA-ED," March 8, 2011. Available at: <https://www.ed.gov/sites/default/files/arpa-ed-background.pdf>
²⁰¹ Rick Hess, "ARPA-ED: A Qualified Thumbs-Up," *Education Week*, March 2, 2012. Available at: http://blogs.edweek.org/edweek/rick_hess_straight_up/2012/03/arpa-ed_a_qualified_thumbs-up.html

Evidence

"I think on the evidence building, we got a lot of things right," said Shelton. "I think that we were smart to arrange for technical assistance to help the grantees structure their evaluations to get the most rigorous, appropriate evidence. There were a number of folks who had consultants, but they found they were putting together a study that would not have served them well in the end."

"Evidence-building is a boutique endeavor," said Dabby. "There are a small number of organizations that do it well. So far, most of the evidence that has been generated under i3 has been specific to the project that they took on. Did the project work? It is a little binary: red light, green light."

"If we really want to democratize evidence the way ESSA envisions, we need different and better evidence," she said. "We still need to know if it works, but also what about it was most impactful and why."

"I think if I had it to do over, I would have spent more of the scale up money on implementation studies, not impact studies," said Shelton. "Because one of the questions is when you move to scale, given the evidence is already high, the real question is: can you scale with fidelity and produce results?"

What about the pace of research? The first grants from i3 were made in 2010, but it was not until the end of the administration that those grants began to produce results.

"How do we speed up the process?" asked Dabby. "On i3 and EIR, you can apply for a three-year grant. It's the same amount of money, so you get more per year if you apply for a shorter grant. But few people do it and the ones that do have all needed extensions."

"We didn't have to wait until now to know that those early grants were showing success," said Shelton. "We could see what was happening in a lot of the interim studies."

"But these are questions you only ask at the beginning," he said. "No one thinks about the most important breakthroughs that started 25 years ago at NIH. What we really need to do is keep going, so there is a constant pipeline of new things that are coming out."

Scale

Scale was a central feature of i3 and it continues to be under EIR. In interviews, there were different ideas about how well this has worked or how it could

be improved.

"I don't see i3 as successfully scaling," said Whitehurst. "Why do you need to scale up Success for All and Teach for America? They are widely-scaled already. The notion that the feds should pay for that doesn't make sense. Spending most of the money to create wider implementation of programs that are already well-established doesn't support innovation."

"What is the theory of action behind scaling?" he asked. "What motivates education administrators writ large? i3 is a bribe. They compete, they agree to do things, they do it. My guess is that most of that goes away once the money isn't on the table anymore."

"You need a different rationale," he said. "One is accountability. It's fine to prime the pump, but if you don't have backup to make the local principals care, then you are spitting into the wind."

"The peak funding was \$650 million. There is only much scale you can accomplish with that," said West. "Scale is a misnomer for describing anything that EIR can do. The right way to think about it is that the scaling grants are part of the evidence building process – testing something in multiple settings."

"What we were trying to do in i3 was innovation, evidence, and scale," said Shelton. "But at a higher level, I think we were trying to create a marketplace for evidence where some people are looking for evidence to make choices and other people are providing evidence to be chosen. By creating that marketplace, the incentives are aligned. If people can keep that kind of framing in mind, regardless of the program, that is the thing that is most important," he said.

"It is ironic that the highest evidence thresholds in the Department are attached to competitive programs of a few hundred million dollars, while billions are going out with no evidence requirements at all," he said. "If even a small portion of Title I dollars prioritized the top two evidence tiers, even just 10 percent, you would have a \$1.4 billion marketplace. That would significantly change the incentives for people seeking Title I funding."

He saw some of that coming from regulations being developed under ESSA, but he urged legislators not to be too restrictive. "The regulations need to be allowed to adapt quickly based on what we learn to continually become more effective," he said. "The notion that we would design something this complicated and be expected to get everything right the first time is flawed."

Even if the right incentives were put in place, he

still saw a need to help states and local school districts.

“Think about how corporates decide whenever they are trying to do something new,” he said. “If they are implementing a new software system in HR, most of the time they will bring in outside organizations who have done it many times before to help with implementation.”

Dabby thought that the Regional Education Laboratories at IES could help states and school districts with this. “They could leverage the i3 and EIR grants,” she said. “That also keeps people in their lane. Supporting states and districts to build and leverage research is part of their role.”

She also thought that state-focused external organizations should provide support to states and districts that want to engage deeply with ESSA’s evidence provisions for formula funds.

Outside the Bubble

The national experts interviewed for this report were brimming with ideas. For these ideas to take hold, however, they must be widely embraced by educators, policymakers, and the public. So far, there is only modest evidence of that.

Is the concept of evidence too esoteric? How can decision makers and the broader public come to understand its importance?

“It is important for us to build evidence,” said Shelton, who has given the topic a lot of thought, “but what is most important is that it is conveyed with actual stories of how that evidence reflects improvements in real people’s lives.”

“We need to move the needle in a visible way, not just in a statistical way,” said Whitehurst. “We need something where Aunt Sarah can see it and get it.”

–

Baltimore is only 40 miles from the halls of power in Washington, DC, but it feels like it is worlds away.

Deep in the heart of one of its toughest neighborhoods is Franklin Square, a K-8 school that is one of the city’s rare gems. The neighborhood suffers high rates of teen pregnancy and recidivism. Boarded up row houses are a common sight. But the story inside the school is very different.

A few years ago, it began working with Success for All as part of its i3 grant. “Our scores were not the best in reading and we were looking for a vehicle that would help our scholars reach success and do it quickly,” said Terry Patton, the school’s hard charging principal.

How well is it working? This time the answer did not come from an evaluation, but from a young woman for whom that answer meant everything.

Tyria is the mother of two children at Franklin Square. She knows too well how tough the city can be because she was homeless once and she lived it. Now she is a school volunteer.

“I like being part of the school,” she said. “Everybody at Franklin is family. If you need a haircut, whatever you need, you can get it from Franklin Square.”

Her two children attended three other schools before they came to Franklin. They were behind when they arrived, but have made enormous progress since then.

“My son is in third grade now. His reading has gotten better. There are so many things they didn’t know,” she said. Her daughter is now a year ahead of grade-level reading. “From where she came from to now is so different.”

So is her life trajectory. “I want her to graduate, go to college, get a good job,” she said softly.

“She says she’s going to be a teacher or a nurse.”

Recommendations

Early results from the Education Innovation and Research (EIR) program (formerly i3) have been generally positive. It has produced 13 projects with evaluations reporting positive impacts and done so at rates that appear to exceed those in other areas of education research. If current rates are sustained, other projects still in the pipeline will produce another 39 evaluations with positive findings over the next few years.

The early scale-up grantees have also been generally successful. All four with final evaluations expanded their programs, two with positive impact findings and two with mixed results. These grantees have also provided insights on what is necessary to scale evidence-based programs effectively.

Despite these early successes, however, the program has several serious shortcomings. EIR could be improved by Congress and the Trump administration with the following supportive changes:

- **EIR Should Be Reworked to Better Find, Support, and Test Groundbreaking Innovations:** To date, finding and supporting truly innovative solutions to the nation's education needs has not been an obvious strength of the i3 program. The new EIR program has taken steps to address this issue by supporting a greater focus on continuous improvement, but more is needed.

One area that needs greater attention is the grantee selection process. While i3 and the new EIR program have had the luxury of selecting from a large number of applicants, it is not clear that the selection process has been well designed to choose grantees with the most promising ideas. The EIR program should review the current process to determine how it could be improved.

Options include reviewing conflict-of-interest policies for peer reviewers to ensure that recognized experts who do not have a direct financial stake are not being unnecessarily

disqualified; inviting applicants who are finalists to make in-person presentations; and adopting an intermediary model similar to that used by the Social Innovation Fund, where external organizations select grantees after conducting substantial due diligence.²⁰² The last of these may require congressional approval.

Finally, interviews for this report found substantial appetite among lower-capacity development grantees who want better connections to, and individualized advice and assistance from, national experts in their respective fields of interest. The program should heighten its attention to this need.

- **EIR Should Support Faster Research:** One major problem with both i3 and the new EIR program is that producing research takes too long. The first grants under i3 were made in 2010 and the final results did not begin to appear in substantial numbers until 2016. While some research will take more time, six years is too long to wait for results in most cases.

Former Obama administration officials say that attempts to speed the process through shorter grant periods have not worked, that the pace of research is a widespread problem that also confronts health and other areas of research, and that this is of less concern now that a pipeline has been built, with new results now expected to be produced every year. There are also tensions between the need to promote true innovation, which can take longer, and the need for quick results.

While somewhat persuasive, these answers too easily dismiss this problem. Much of the delay is due to funds being used to expand programs at the same time that they are being evaluated, creating delays as initiatives are set up in new schools and new staff are hired, trained, and gain necessary experience before the evaluation can begin. In some cases, this is necessary to

²⁰² Social Innovation Research Center, "Social Innovation Fund: Early Results Are Promising," June 20, 2015, p.p. 19, 34-38. Available at: <http://www.socialinnovationcenter.org/wp->

[content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf](http://www.socialinnovationcenter.org/wp-content/uploads/2015/07/Social_Innovation_Fund-2015-06-30.pdf)

provide insights on scaling or to ensure appropriate sample sizes for multi-site evaluations, but this is not always true.

To provide faster research results, the program should support more evaluations of existing programs in its early-phase and mid-phase grants, without further expansion. It should also support greater use of lower-cost, short-duration grants like those that have been funded by the Institute of Education Sciences.²⁰³

- **To Better Support Scaling of Evidence-based Programs and Practices, EIR Should Be Better Connected to Other Publicly-funded Programs:** Any strategy for successfully scaling evidence-based programs or practices must address two critical components: supply and demand. In education, the supply side is being addressed by EIR, the Institute of Education Sciences (IES), and other federal, state, local, and private entities.

Demand for evidence-based education programs is still limited, but it took a significant step forward under ESSA. Two important demand drivers are its reworked state accountability measures²⁰⁴ and new evidence requirements that have been built into a variety of formula-funded and competitive grant programs.²⁰⁵

These provisions, which are still being rolled out, will help. But this report and earlier research strongly suggest that successfully scaling and adopting evidence-based programs may require the support of intermediaries with deep experience that can help states and local school districts put these programs in place. The need for such intermediaries may be greatest in schools that are low-performing.

At the federal level, the Regional Education Laboratories (RELs) play a major role in disseminating and supporting the adoption of evidence-based practices. However, these laboratories are primarily responsive to the needs of states. They are neutral with respect to developers that are disseminating their own

evidence-based models. Given their location within IES, which is protective of its reputation for impartiality, the RELs would not be an appropriate vehicle for supporting such intermediaries.

EIR's continued ability to support these intermediaries may be crucial. However, changes may be needed to address some of the program's limitations.

The most important of these is budgetary. With an annual budget of \$120 million, EIR's ability to support such intermediaries is limited. Any substantial expansion of its budget by the incoming administration also seems unlikely.

One promising, although not widely noted, development came when Congress authorized the use of federal, state, and local government funds for matching purposes under EIR. This strategy should be taken further.

EIR should provide competitive preferences for applicants that can leverage other federal, state, and local funds in EIR's expansion and mid-phase grants. Aligning EIR with other programs in this way would not only multiply the reach of EIR's limited budgetary resources, it would also help infuse evidence into these other federal, state, and local programs.

- **To Avoid the Appearance of Politicization, Congress and the Trump Administration Should Consider Moving EIR to the Institute of Education Sciences:** There is disagreement among national experts over the proper placement of the EIR program at the Department of Education, with some saying it should be moved to IES and others saying should stay at the Office of Innovation and Improvement (OII).

Moving the program to IES would provide it with legal protections under the Education Sciences Reform Act that would shield it from perceived or actual politicization. It would also provide it with better access to the research-related expertise of that agency.

²⁰³ Institute of Education Sciences, "Low-Cost, Short-Duration Evaluations: Helping States and School Districts Make Evidence-based Decisions," May 23, 2016. Available at: <https://ies.ed.gov/blogs/research/post/low-cost-short-duration-evaluations-helping-states-and-school-districts-make-evidence-based-decisions>

²⁰⁴ Alyson Klein, "Final ESSA Accountability Rules Boost State Flexibility in Key Areas," *Education Week*, November 28, 2016. Available at: [http://blogs.edweek.org/edweek/campaign-k-](http://blogs.edweek.org/edweek/campaign-k-12/2016/11/ed_dept_releases_final_account.html)

[12/2016/11/ed_dept_releases_final_account.html](http://blogs.edweek.org/edweek/campaign-k-12/2016/11/ed_dept_releases_final_account.html); and ASCD, "ESSA and Accountability: Frequently Asked Questions," May 2016. Available at: http://www.ascd.org/ASCD/pdf/siteASCD/policy/ESSA-Accountability-FAQ_May112016.pdf

²⁰⁵ Patrick Lester, "K-12 Education Bill Advances Evidence-based Policy, Replaces i3," Social Innovation Research Center, December 7, 2015. Available at: <http://www.socialinnovationcenter.org/?p=1806>

Keeping it at OII would make it more responsive to the policy interests of the incoming administration, including in school choice, accountability, or other issues. Keeping it at OII would also not affect reviews by the What Works Clearinghouse, which is part of IES, has the same protections for independence as IES, and is the final arbiter of evaluation quality. Finally, staying at OII could also help ensure that the program retains a practitioner orientation, including providing capacity-building to nonprofits and schools and focusing on issues of most immediate concern to them.

This report takes no position on this issue other than to acknowledge the tradeoffs and to recommend that Congress and the new administration consider the matter.

- **To Better Promote Bottom Up Innovation, EIR Should Reconsider Its Use of Program Priorities and Competitive Preferences:** One criticism leveled at the i3 program is that the Obama administration was too heavy-handed in its use of priorities and competitive preferences during the grantee selection process, potentially discouraging projects that were not aligned with its education agenda. This criticism was part of a larger, bipartisan backlash against a federal role in K-12 education that some believed had become too intrusive.

When Congress enacted ESSA, advocates who supported EIR said it would discontinue the use of these federally-determined priorities.²⁰⁶ Some were later surprised when the Department of Education retained their use when it announced the availability of new grants under the program, apparently drawing on other regulatory authority.²⁰⁷ In interviews, OII staff have indicated that there are benefits that come from choosing grantees with similar areas of focus, including the creation of communities of practice.

The Trump administration will likely revisit this decision. In doing so, however, it should consider other ways to ensure that new projects are focused on gaps in education research. So far, finding and testing truly innovative ideas seems to be a weakness of the program. Further distancing the program from the guidance of national experts, who are aware of gaps in research and may be well positioned to identify true innovations, could make this worse – possibly resulting in multiple replications of existing evidence-based practices, with little that is truly new or innovative.

Addressing this need does not necessarily require the use of top-down absolute priorities and competitive preferences. However, it may provide further reason to consider changing the program's selection process, which is discussed in the first recommendation on innovation.

- **EIR Should Support Greater Transparency in Its Evaluation Results:** Evaluation results, even when they are poor, can provide important insights to policymakers and practitioners.²⁰⁸ The EIR program currently posts links to final evaluations on its web site,²⁰⁹ but these links are not easy to find and they are vulnerable to being lost when organizations make routine changes to their web pages. Moreover, while What Works Clearinghouse (WWC) study reviews are available for some of these evaluations, links to them have not been posted on the EIR / i3 web site and they can be difficult to identify and locate on the WWC web site.

This should change. The Department should ensure that all final i3 and EIR evaluations are posted at the Education Resources Information Center (ERIC).²¹⁰ Links to these reports and to WWC study reviews should also be prominently displayed on the EIR web site, preferably alongside the grant application materials that are already posted there.²¹¹

²⁰⁶ Patrick Lester, "K-12 Education Bill Advances Evidence-based Policy, Replaces i3," Social Innovation Research Center, December 7, 2015. Available at:

<http://www.socialinnovationcenter.org/?p=1806>

²⁰⁷ Patrick Lester, "ED Announces First Round of Grants Under i3's Replacement," Social Innovation Research Center, December 15, 2016. Available at:

<http://www.socialinnovationcenter.org/?p=2430>

²⁰⁸ Lucy Goodchild van Hilten, "Why It's Time to Publish Research 'Failures'," May 5, 2015. Available at:

<http://www.elsevier.com/connect/scientists-we-want-your-negative-results-too>

²⁰⁹ See <https://www2.ed.gov/programs/innovation/awards.html>

²¹⁰ By comparison, IES has a free public access policy that requires all IES grantees to submit their published, peer-reviewed work to ERIC within one year of publication. See: <https://ies.ed.gov/funding/researchaccess.asp>

²¹¹ In an interview on January 17, 2017, OII staff said that submission to ERIC is a new requirement under EIR and that they plan to put links to WWC study reviews on their web site.

- **To Improve the Chances of Replicating Successful Programs, EIR Should Support the Development of Higher Quality Fidelity Measures:** To date, the Department has provided technical assistance on evaluations through Abt Associates. This support appears to be well-regarded by the grantees and it seems to have played a critical role in ensuring that most project evaluations are well positioned to receive strong What Works Clearinghouse ratings.

One critical weakness, however, is the quality of project fidelity measures. High quality fidelity

measures can provide an early-warning system for poor performance, allowing needed course corrections. They can also provide important insights if a project fails to produce positive impact or, if evaluation findings are positive, provide a basis for identifying core program components that should be replicated.

This report's review of the final i3 evaluations found that their fidelity measures often seemed pro-forma and poorly constructed. These measures need more attention from Abt Associates and IES.

Appendix A: Project Summaries

Summary results for all four 2010 scale-up grants and the 13 projects rated in this report as having positive impact can be found below (15 projects total). Validation and development grant projects with mixed results (i.e., positive results on at least one, but fewer than half of important impact measures) are not summarized. Links to all 44 evaluations, including those with mixed results or no impact, can be found in Appendix B.

Evaluation results are based on findings as stated in the final evaluations. Where available, the results reported here also incorporate What Works Clearinghouse study reviews, with links provided to the review. Otherwise, they do not reflect a detailed review of the underlying evaluation methodology and results, a significant undertaking normally performed by an evidence clearinghouse.

ASSET Inc.: 2010 Validation – Standards and Assessments (\$20,230,572)

- Title = ASSET Regional Professional Development Centers for Advancing STEM Education
- Intervention = STEM professional development for 24 high-needs elementary schools (grades K-6) in rural Pennsylvania. The initiative's professional development activities supported a science curriculum aligned with state standards. Components included summer professional development trainings, coaching, and the creation of a Professional Learning Community.
- Evaluation = The project was evaluated with a quasi-experimental design (QED) evaluation that compared academic results for enrolled students on state math and science test scores to those in two comparison groups of schools, one of which had participated in other statewide science initiatives and one of which had not.
- Results = The study showed statistically significant improvements for enrolled students in fourth grade science and third grade math when compared to schools that had not participated in other statewide science initiatives. No statistically significant results were found when compared to other schools that had participated in other statewide science initiatives.
- Final Report = https://assetinc.org/files/public/asset_pd_impact_on_achievement.pdf

Bellevue School District: 2010 Development – Standards and Assessments (\$4,149,778)

- Title = Re-imagining Career and College Readiness
- Intervention = A program that redesigned courses in a single high school (Sammamish High School) in Bellevue, WA that used a Problem-Based Learning (PBL) strategy. The program focused on students from groups that have been traditionally underrepresented in STEM fields. It combined an intensive summer program with mentoring from professionals in various fields, including STEM, with an intensive opt-in summer program serving a subset of students.
- Evaluation = The mixed method evaluation combined significant qualitative research with a quasi-experimental design (QED) study that compared students enrolled in the program to a matched cohort of other students in the school that were not. It examined AP exam scores and results of a campus readiness assessment.
- Results = The study indicated that students enrolled in the program had higher AP scores in technology, engineering, and mathematics and improved college and career readiness, as determined by results on the Educational Policy Improvement Center's CampusReady survey.
- Final Report = <http://www.bsd405.org/wp-content/uploads/2016/03/Sammamish-i3-Grant-Findings-Report.pdf>

Boys and Girls Clubs of Milwaukee: 2010 Development – Low-performing / School Turnaround (\$4,142,965)

- Title = Milwaukee Community Literacy Project (SPARK)
- Intervention = Two-year holistic program that features in-school tutoring and family engagement for 300 K-3 students at seven Milwaukee schools intended to help them reach reading level by the beginning of 4th grade. The program includes research-based one-on-one in-school tutoring by trained tutors, after-school supplementary reading sessions, and regular contact with parents and home visits to increase parents' skills in supporting their child's literacy. The initiative had been under development since 2005.
- Evaluation = The program was evaluated with a randomized controlled trial (RCT) study that compared outcomes for enrolled students to a control group of other students that received "business as usual"

reading instruction from Milwaukee public schools.

- **Results** = The study showed improvements in reading achievement, literacy, and school attendance.
- **Final Report** = <https://uwm.edu/education/research/socially-responsible-evaluation-in-education/milw-community-literacy-spark/>
- **What Works Clearinghouse Review** = <https://ies.ed.gov/ncee/wwc/Study/32028>

Children’s Literacy Initiative: 2010 Validation – Teacher and Principal Effectiveness (\$21,726,296)

- **Title** = Model Classroom Project
- **Intervention** = Provided literacy instruction for K-3 teachers to help them implement “model classrooms.” The initiative provided intensive coaching and support to one teacher per grade to prepare him or her to help colleagues also use best practices.
- **Evaluation** = The study was a randomized controlled trial (RCT) conducted in 78 schools across four school districts in Chicago, Newark, Camden, and Philadelphia. Participating schools were randomly assigned to the treatment or control group. Treatment schools received services for students in grades K-2 in the first three years of the grant.
Teachers at the 39 schools in the control group received only the professional development normally provided by their district. As an incentive for participation, schools in the control group received a \$4,000 school library. Control group schools began receiving services after the study in the final two years of the grant.
- **Results** = The study found that the program produces substantial effects on teachers’ classroom environment and literacy practice. After three years in the program, second graders scored statistically higher in a study-administered assessment of reading achievement, the Group Reading Assessment and Diagnostic Evaluation (GRADE).
- **Final Report** = <http://www.cli.org/impact/i3-air/>
- **What Works Clearinghouse Review** = <https://ies.ed.gov/ncee/wwc/Study/81569>

Fresno County Office of Education: 2011 Development – Standards and Assessments (\$3,000,000)

- **Title** = Expository Reading and Writing Course
- **Intervention** = A reading and writing course for high school seniors intended to help them become better prepared for college. The course was designed by California State University (CSU) to reduce the need for remediation in English for first-year college students and is aligned with Common Core state standards in reading and writing. The course was administered to students in 24 high schools in California.
- **Evaluation** = The study was a quasi-experimental design (QED) that compared the reading and writing skills of students enrolled in the course to a matched comparison group of students who took a different English class.
- **Results** = Enrolled students scored higher than the comparison students on the English Placement Test. The difference was statistically significant at the 1 percent level.
- **Final Report** = <https://www.wested.org/resources/evaluation-of-expository-reading-writing-course/>
- **What Works Clearinghouse Review** = <http://ies.ed.gov/ncee/wwc/Study/32029>

Iredell-Statesville Schools: 2010 Development – Teacher and Principal Effectiveness (\$4,999,036)

- **Title** = COMPASS: Collaborative Organizational Model to Promote Aligned Support Structures
- **Intervention** = This school district in North Carolina tested a professional development initiative for middle school teachers that helped them identify students that are struggling and address their individual academic needs. The initiative was implemented in all 21 of the district’s schools.
- **Evaluation** = The evaluation used a quasi-experimental (QED) design that compared student scores on the End-of-Grade state reading test in 21 district schools in grades 3-8 to students in schools in neighboring school districts that had been matched according to demographic characteristics.
- **Results** = The study found a positive effect on student reading achievement.
- **Final Report** = <http://iss.schoolwires.com/domain/5903>

KIPP Foundation: 2010 Scale-up –Teacher and Principal Effectiveness (\$50,000,000)

- Title = Scaling-Up KIPP's Effective Leadership Development Model
- Intervention = This grant supported the scale up of KIPP, a national network of charter schools. Under the grant, KIPP was expanded to additional elementary, middle, and high schools.
- Evaluation = The study combines a randomized controlled trial (RCT) and quasi-experimental design (QED) that examined impacts on student achievement at 8 elementary, 43 middle, and 18 high schools in 20 cities. The RCT offered students admission to KIPP schools by lottery. Students not offered admission through the lottery enroll at other charter, private, or traditional public preschools or elementary schools and are included in the control group. The study used data from study-administered student achievement tests; state assessments in math, English/language arts (ELA), science, and social studies; and student and parent surveys.
- Results = The study found that KIPP elementary schools have a positive impact on student reading and math achievement and that KIPP middle schools have positive impacts in math, reading, science, and social studies. It also found positive impacts on parent satisfaction with their child's school. However, it found no impacts on student motivation, engagement, educational aspirations, or behavior.
- Final Report = <http://www.kipp.org/results/independent-reports/#mathematica-2015-report>

Niswonger Foundation: 2010 Validation – Standards and Assessments (\$17,751,044)

- Title = Northeast Tennessee Consortium (NETCO)
- Intervention = A college- and career-readiness program for 15 school districts in rural northeastern Tennessee that consists of six components: (1) management and communication, (2) promoting a college-going culture, (3) quality of instruction, (4) distance and online technology, (5) college-level courses, and (6) resources and services to expand and sustain program capacity.
- Evaluation = The evaluation was a quasi-experimental design (QED) that compared results between participating schools and comparison schools that were identified through propensity score modeling.
- Results = Students in participating schools had higher ACT scores, were more likely to participate in Advanced Placement (AP) courses, score a 3 or higher on an AP exam, enroll in college, and persist in college than students in matched comparison schools.
- Final Report = https://www.cna.org/cna_files/pdf/Final%20Findings%20from%20Impact%20and%20Implementation%20Analyses%20of%20the%20Northeast.pdf

Search Institute / BARR: 2010 Development – Low-performing Schools / School Turnaround (\$4,999,711)

- Title = Building Assets-Reducing Risks (BARR) Turnaround Project
- Intervention = Project designed to increase high school graduation and college enrollment rates by providing supports for students in 9th grade. The program was implemented in three schools, one in suburban Los Angeles and two in rural Maine.
The program organizes students into cohorts of 30 who take courses together in math, English, and science or social studies. It also provides professional development for teachers, counselors, and administrators and holds regular meetings of cohort teacher teams that include addressing persistently low-performing students. It also includes a family engagement component.
- Evaluation = The Los Angeles program was studied using a randomized controlled trial (RCT) that randomly assigned students to the program. The two Maine schools were not part of the RCT.
- Results = Enrolled students earned more core credits, obtained better grades, experienced lower course failure rates, and earned higher test scores in reading and mathematics than students not enrolled in the program.
- Final Report = <http://www.barrcenter.org/results>
- What Works Clearinghouse Review = <https://ies.ed.gov/ncee/wwc/Study/132>

Success for All: 2010 Scale-up – Low-performing Schools / School Turnaround (\$49,285,513)

- Title = Scale-Up and Evaluation of Success for All in Struggling Elementary Schools
- Intervention = This grant scaled up the SFA whole-school turnaround model in elementary schools. Key components of SFA include an extensive reading program for students in kindergarten through grade 6, job-embedded professional development and coaching, collaborative performance monitoring, curriculum resources, and strategies for addressing school-wide issues such as low attendance, parental involvement, school culture, family needs, and health issues.

- Evaluation = The initiative was evaluated with a randomized controlled trial (RCT) in which 37 high-poverty elementary schools across the United States were randomly assigned to receive SFA or not.
- Results = The evaluation found positive effects on student phonics, particularly students with low pre-literacy skills, but found no effects on reading comprehension, special education designations, or rates at which students were held back to repeat a grade.
- Final Report = <http://www.mdrc.org/publication/scaling-success-all-model-school-reform>
- What Works Clearinghouse Review = <https://ies.ed.gov/ncee/wwc/Study/32024>

Teach for America: 2010 Scale-up – Teacher and Principal Effectiveness (\$50,000,000)

- Title = Scaling Teach for America
- Intervention = This grant supported a scale up of Teach for America, a nonprofit program that recruits college graduates and professionals with strong academic backgrounds and leadership experience to teach for two years in high-needs schools. Participants typically have no formal training in education and participate in an intensive five-week training program before beginning their first teaching job. TFA then provides ongoing training and support throughout their two-year commitment.
- Evaluation = The initiative was evaluated with a randomized controlled trial (RCT) that compared the effectiveness of TFA elementary school teachers with incumbent teachers. Students in 36 schools were randomly assigned to either the TFA teachers or to other teachers in those schools who had been certified through traditional means. Results were based on student achievement on end-of-year reading and math test scores from the 2012–2013 school year.
- Results = The evaluation found a statistically significant positive impact on student reading achievement for TFA teachers in lower elementary grades (K through grade 2) and a marginally significant positive impact on student math achievement for TFA teachers in grades 1 and 2. It otherwise found comparable results for the control and treatment groups, a finding that earns it a “mixed impact” rating in this report. However, this rating comes with an important caveat. According to the study, TFA teachers had an average of 1.7 years of experience compared with 13.6 years among the comparison teachers, so the finding of similar results for the program’s new teachers when compared to substantially more experienced incumbent teachers still suggests a positive result.
- Final Report = <http://www.mathematica-mpr.com/news/assessing-the-effectiveness-of-the-teach-for-america-i3-scale-up>
- What Works Clearinghouse Review = <https://ies.ed.gov/ncee/wwc/Study/1137>

The Ohio State University: 2010 Scale-up – Low-performing Schools / School Turnaround (\$45,593,146)

- Title = Reading Recovery: Scaling Up What Works
- Intervention = This initiative scaled up the evidence-based Reading Recovery program for struggling first grade students. Reading Recovery is an intensive intervention that includes 12- to 20-weeks of daily, one-to-one Reading Recovery lessons provided by a trained teacher as a supplement to regular classroom literacy instruction. The program was implemented in over 1,300 schools.
- Evaluation = The impact evaluation includes a multi-site randomized controlled trial (RCT) for estimating immediate impacts, a regression discontinuity study (RD) for estimating long term impacts, and a mixed-methods study of program implementation under the i3 scale-up. The RCT matched students within schools based on pretest scores and randomly assigned them to the treatment or control groups. Students in the control group received regular classroom literacy instruction as well as any interventions normally provided to low-performing 1st-grade readers in their schools.
- Results = The RCT revealed medium to large impacts across all outcome measures based on scores on the Iowa Test of Basic Skills (ITBS) Reading Total assessment, the ITBS Reading Comprehension and Reading Words subtests, and on the Observation Survey of Early Literacy Assessment (OS). The evaluation also found similar results in two subgroups of interest: English Language Learners and students in rural schools. The regression discontinuity study largely replicated the RCT findings. The implementation study revealed strong fidelity to the model and that teachers were properly trained.
- Final Report = <http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>
- What Works Clearinghouse Review = <https://ies.ed.gov/ncee/wwc/Study/32027>

The Studio in a School: 2010 Development – Standards and Assessments (\$4,372,798)

- Title = Arts Achieve: Impacting Student Success in the Arts
- Intervention = This program focused on improving arts achievement through the creation of validated, open access educational resources and assessments. The first year of the project was spent developing benchmark assessments. The program was then implemented in participating New York City schools, including workshops, in-class support, teacher peer observations, technology, and other supports.
The project brought together the New York City Department of Education's (NYC DOE) Office of Arts and Special Projects and five cultural arts organizations in NYC: Studio in a School (lead partner; visual arts), ArtsConnection (theater), the Weill Music Institute at Carnegie Hall (music); the Dance Education Laboratory at the 92nd St. Y (dance), and the Cooper Hewitt Smithsonian Design Museum (technology).
- Evaluation = The program was evaluated in a randomized controlled trial (RCT) that compared results for students in schools that met basic eligibility requirements, were recruited, and subsequently randomly assigned to either the treatment or control group.
- Results = In each year of implementation, students in the treatment schools made greater gains in arts achievement than students in the control schools. The Year 1 effect size was 0.28, the Year 2 effect size was 0.20, and the Year 3 effect size was 0.09.
- Final Report = http://www.studioinaschool.org/uploads/6/6/6/4/6664061/arts_achieve_-_summary_of_findings_082715.pdf

Utah State University: 2010 Validation – Low-performing Schools / School Turnaround (\$15,282,720)

- Title = New Mexico StartSmart K-3 Plus
- Intervention = The program is intended to improve kindergarten readiness and academic achievement for K-3 students in New Mexico. Primary program components include a 25-day summer program for students, professional development in literacy for participating teachers, and parent outreach.
- Evaluation = The program was evaluated in a randomized controlled trial (RCT) study that compared results for students who were enrolled in the program to those in a control group of students that only received regular school year services.
- Results = Participating students who attended the program before kindergarten performed better than the control group on tests of vocabulary, reading, writing, and math, but not social skills or receptive language. By the start of 3th grade, students who were starting their fourth year of the program performed better than the control group in reading, math, and writing, but not vocabulary, social skills, or receptive language.
- Final Report = <http://startsmark3plus.org/>

University of Missouri: 2010 Validation – Standards and Assessments (\$12,277,674)

- Title = eMINTS Professional Development on Student and Teacher Outcomes
- Intervention = Comprehensive professional development program that provides 240 hours of training over two years to design high-quality inquiry-based lesson plans, implement inquiry-based learning strategies, build community among teachers and students, and integrate technology into classroom instruction.
- Evaluation = The RCT-based study randomized 60 high-poverty rural Missouri middle schools, with one group of schools receiving the traditional eMINTS program, the second receiving eMINTS plus a third year of professional development using Intel Teach Program courses, and the third acting as a control group and receiving business-as-usual services.
- Results = The study showed statistically significant improvements for students in mathematics, but not English, for both traditional eMINTS and eMINTS plus. Teachers also showed statistically higher scores in inquiry-based practices and technology integration. Finally, eMINTS plus, but not eMINTS, scored better on high-quality lesson design.
- Final Report = http://emints.org/wp-content/uploads/2014/05/eMINTS-Research-Findings-Summary_updated-04.15.2015.pdf
- What Works Clearinghouse Review = <https://ies.ed.gov/ncee/wwc/Study/81457>

Appendix B: Links to Final Evaluations

Links to the final evaluations for all 44 projects included in this analysis are below. Projects appended with an asterisk (*) are summarized in Appendix A. Other projects that are listed here, but not appended with an asterisk, may have mixed results (i.e., positive results on at least one, but fewer than half of important impact measures).

Alliance for College Ready Schools: 2010 Development - Standards and Assessments
CollegeYes
<https://bpe.egnyte.com/dl/ssajC1W4Fx>

American Federation of Teachers: 2010 Development – Teacher and Principal Effectiveness
Excellence in Teaching and Learning Consortium
https://i3community.ed.gov/system/files/resource_files/2016/e3tl_evaluation_final_report.pdf

AppleTree Institute: 2010 Development – Data/Data Driven Instruction
Every Child Ready
<http://www.appletreeinstitute.org/wp-content/uploads/2016/08/Evaluation-Report-Submitted-to-Abt-June-2015.pdf>

Aspire Public Schools: 2011 Development – Teacher and Principal Effectiveness
Transforming Teacher Talent (t3)
<https://www.empiricaeducation.com/pdfs/AspireFR.pdf>

ASSET Inc.: 2010 Validation – Standards and Assessments *
ASSET Regional Professional Development Centers for Advancing STEM Education
https://assetinc.org/files/public/asset_pd_impact_on_achievement.pdf

Baltimore City Public Schools: 2011 Development – Promoting STEM Education
Middle School STEM Summer Learning Program
<http://baltimore-berc.org/the-baltimore-city-schools-middle-school-stem-summer-program-with-vex-robotics/>

Bay State Reading Institute: 2010 Development – Data/Data Driven Instruction
Bay State Reading Institute (BSRI)
http://www.baystareading.org/wp-content/uploads/2015/12/20152508_SchoolWorks_BSRI_Y5_FinalReport.pdf

Beaverton School District 48J: 2010 Development – Standards and Assessments
Arts for Learning Lessons Project (A4L)
https://www.beaverton.k12.or.us/depts/tchlrn/lts/arts4lrng/A4L/2015-2016/Student_Impact_Findings.pdf

Bellevue School District: 2010 Development – Standards and Assessments *
Re-imagining Career and College Readiness
<http://www.bsd405.org/wp-content/uploads/2016/03/Sammamish-i3-Grant-Findings-Report.pdf>

Board of Education for New York City: 2010 Development – Data/Data Driven Instruction
School of One
<http://www.wexford.org/equity-projects>

Boston Plan for Excellence in the Public Schools: 2010 Development – Teacher and Principal Effectiveness
Boston Teacher Residency
<https://bpe.egnyte.com/dl/ssajC1W4Fx>

Boys and Girls Clubs of Milwaukee: 2010 Development – Low-performing Schools / School Turnaround *
Milwaukee Community Literacy Project (SPARK)
<https://uwm.edu/education/research/socially-responsible-evaluation-in-education/milw-community-literacy-spark/>

California Education Roundtable: 2010 Development – Standards and Assessments
STEM Learning Opportunities Providing Equity (SLOPE)
http://arches-cal.org/wp-content/uploads/2016/06/WestEd_Final-Evaluation-Report_SLOPE-i3_09-30-2015.pdf

Central Falls School District: 2012 Development – Parent and Family Engagement
We Are A Village
<http://annenberginstitute.org/sites/default/files/product/873/files/i3Report2016Web.pdf>

Children’s Literacy Initiative: 2010 Validation – Teacher and Principal Effectiveness *
Model Classroom Project
<http://www.cli.org/impact/i3-air/>

Corona-Norco Unified School District: 2010 Development – Standards and Assessments
Write Up!
<http://www.cnusd.k12.ca.us/i3>

Exploratorium - Institute for Inquiry: 2010 Development – Teacher and Principal Effectiveness
Exploratorium Institute for Inquiry
<http://www.exploratorium.edu/education/ifi/inquiry-and-eld/educators-guide/project-studies>

Education Connection: 2010 Development – Standards and Assessments
Science, Technology, Engineering, and Math Education for the 21st Century (STEM21)
<http://www.skills21.org/about/research>

Forsyth County Schools: 2010 Development Grant – Data/Data Driven Instruction
EngageME-P.L.E.A.S.E.
http://www.forsyth.k12.ga.us/cms/lib3/GA01000373/Centricity/Domain/73/i3_Forsyth_Impact_Report.pdf

Fresno County Office of Education: 2011 Development – Standards and Assessments *
Expository Reading and Writing Course
<https://www.wested.org/resources/evaluation-of-expository-reading-writing-course/>

George Mason University: 2010 Validation – Teacher and Principal Effectiveness
Virginia Initiative for Science Teaching and Achievement (VISTA)
<http://vista.gmu.edu/news-and-research/published-studies>

IDEA Public Schools: 2010 Development – Teacher and Principal Effectiveness
Rio Grande Valley Center for Teaching and Leading Excellence
<http://www.sri.com/work/publications/evaluation-rio-grande-valley-center-teaching-and-leading-excellence-final-report>

Iredell-Statesville Schools: 2010 Development – Teacher and Principal Effectiveness *
COMPASS: Collaborative Organizational Model to Promote Aligned Support Structures
<https://sites.ed.gov/oii/2014/06/iredell-statesville-schools-find-a-boost-from-investing-in-innovation/>

Jefferson County Board of Education: 2010 Development – Low-performing Schools/School Turnaround
Making Time for What Matters
<https://eric.ed.gov/?q=Making+time+for+what+matters+most&id=ED562043>

KIPP Foundation: 2010 Scale-up –Teacher and Principal Effectiveness *
Scaling-Up KIPP's Effective Leadership Development Model
<http://www.kipp.org/results/independent-reports/#mathematica-2015-report>

National Forum to Accelerate Middle-Grades Reform: 2010 Development – School Turnaround
Schools To Watch (STW) Transformation Network
<http://middlegradesforum.org/wp-content/uploads/2016/02/i3-STW-Project-Final-Evaluation-Report-CPRD.pdf>

Niswonger Foundation: 2010 Validation – Standards and Assessments *
Northeast Tennessee Consortium (NETCO)
https://www.cna.org/cna_files/pdf/Final%20Findings%20from%20Impact%20and%20Implementation%20Analyses%20of%20the%20Northeast.pdf

Ounce of Prevention Fund: 2010 Development – Standards and Assessments
Ounce Professional Development Initiative (PDI)
<http://www.theounce.org/what-we-do/research/programs/Investing-In-Innovation>

Parents as Teachers: 2010 Validation – Low-performing Schools/School Turnaround
Improving Educational Outcomes for American Indian Children
http://www.parentsasteachers.org/images/stories/graphics/babyface_report_pat_2016.pdf

Plymouth Public Schools: 2010 Development Grant – Standards and Assessments
New England Network for Personalization and Performance (NETWORK)
<http://i3.cssr.us/sites/default/files/NENPP%20report%20-%20Network%20Final%20Questions.pdf>

Saint Vrain Valley School District: 2010 Development – Low-performing Schools/School Turnaround
St. Vrain Valley School District i3 Project
<http://svvdsd.org/about/departments/investing-innovation-grant-i3/i3-final-report>

School Board of Miami-Dade County: 2010 Development – Teacher and Principal Effectiveness
Florida Master Teacher Initiative
http://earlychildhood.dadeschools.net/pdfs16/FMTI_final_rpt.pdf

School District No. 1/Denver: 2010 Validation – Teacher and Principal Effectiveness
Collaborative Strategic Reading Colorado
<http://curriculum.dpsk12.org/collaborative-strategic-reading/>

Search Institute / BARR: 2010 Development – Low-performing Schools / School Turnaround *
Building Assets-Reducing Risks (BARR) Turnaround Project
<http://www.barrcenter.org/results>

Smithsonian Institution: 2010 Validation – Standards and Assessments
LASER Model: A Systemic and Sustainable Approach for Achieving High Standards in Science Education
https://ssec.si.edu/sites/default/files/Zoblotsky_etal_2016_Smithsonian_LASER_i3_Validation_Report_FINAL_09_01_16.pdf

Success for All: 2010 Scale-up – Low-performing Schools / School Turnaround *
Scale-Up and Evaluation of Success for All in Struggling Elementary Schools
<http://www.mdrc.org/publication/scaling-success-all-model-school-reform>

Take Stock in Children: 2010 Development – Data/Data Driven Instruction
Facilitating Long-Term Improvements in Graduation and Higher Education for Tomorrow
<http://www.takestockinchildren.org/what-we-do/innovations>

Teach for America: 2010 Scale-up – Teacher and Principal Effectiveness *
Scaling Teach for America
<http://www.mathematica-mpr.com/news/assessing-the-effectiveness-of-the-teach-for-america-i3-scale-up>

The Achievement Network LTD: 2010 Development – Data/Data Driven Instruction
Expanding the Achievement Network Model
<http://cepr.harvard.edu/achievement-network-evaluation>

The Ohio State University: 2010 Scale-up – Low-performing Schools / School Turnaround *
Reading Recovery: Scaling Up What Works
<http://www.cpre.org/reading-recovery-evaluation-four-year-i3-scale>

The Studio in a School: 2010 Development – Standards and Assessments *
Arts Achieve: Impacting Student Success in the Arts
http://www.studioinaschool.org/uploads/6/6/6/4/6664061/arts_achieve_-_summary_of_findings_082715.pdf

University of Missouri: 2010 Validation – Standards and Assessments *
eMINTS Professional Development on Student and Teacher Outcomes
<http://emints.org/impact/>

Utah State University: 2010 Validation – Low-performing Schools / School Turnaround *
New Mexico StartSmart K-3 Plus
<http://startsmark3plus.org/>

WestEd: 2010 Validation – Standards and Assessments
Reading Apprenticeship Improving Secondary Education (RAISE)
<http://empiricaeducation.com/blog/reading-apprenticeship-i3-implementation.html>

* Summarized in Appendix A.

— o —

About the Social Innovation Research Center: The Social Innovation Research Center (SIRC) is a nonpartisan nonprofit research organization focused on social innovation and performance management for nonprofits and public agencies. More information about SIRC is available on the organization's web site at <http://www.socialinnovationcenter.org>.

This report was developed with the generous financial support of the **Laura and John Arnold Foundation**. The opinions expressed in this report are the author's and do not necessarily represent the view of the foundation.

SIRC Series on Evidence in Education

Evidence-Based Comprehensive School Improvement

How Using Proven Models and Practices Could Overcome Decades of Failure
March 26, 2018

<http://socialinnovationcenter.org/wp-content/uploads/2018/03/CSI-turnarounds.pdf>

Building and Using Evidence in Charter Schools

How Charter Schools Could Become Innovation Laboratories for K-12 Education
March 26, 2018

<http://socialinnovationcenter.org/wp-content/uploads/2018/03/charterschools.pdf>

Investing in Innovation (i3)

Strong Start on Evaluation and Scale, But Greater Focus Needed on Innovation
January 19, 2017

<http://socialinnovationcenter.org/wp-content/uploads/2017/01/SIRC-i3-report.pdf>