



WHAT DOES IT TAKE TO SCALE UP INNOVATIONS?

AN EXAMINATION OF TEACH FOR AMERICA,
THE HARLEM CHILDREN'S ZONE,
AND THE KNOWLEDGE IS POWER PROGRAM

Ben Levin

University of Toronto

March 2013

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

This is one of a series of briefs made possible in part by funding from
The Great Lakes Center for Education Research and Practice.



<http://www.greatlakescenter.org>
GreatLakesCenter@greatlakescenter.org



Kevin Welner
Project Director

Don Weitzman
Academic Editor

William Mathis
Managing Director

Erik Gunn
Managing Editor

Briefs published by the National Education Policy Center (NEPC) are blind peer-reviewed by members of the Editorial Review Board. Visit <http://nepc.colorado.edu> to find all of these briefs. For information on the editorial board and its members, visit: <http://nepc.colorado.edu/editorial-board>.

Publishing Director: **Alex Molnar**

Suggested Citation:

Levin, B (2013). *What Does It Take to Scale Up Innovations? An examination of Teach for America, the Harlem Children's Zone, and the Knowledge Is Power Program*. Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/publication/scaling-up-innovations>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

WHAT DOES IT TAKE TO SCALE UP INNOVATIONS? AN EXAMINATION OF TEACH FOR AMERICA, THE HARLEM CHILDREN'S ZONE, AND THE KNOWLEDGE IS POWER PROGRAM

Ben Levin, University of Toronto

Executive Summary

Various education innovations are often proposed as solutions to the problems of education in the United States. Moving an innovation from a few schools to a great many, so it can have a regional or national impact, is very challenging, however.

This brief discusses the problem of scaling innovations in education in the United States so that they can serve very large numbers of students. It begins with a general discussion of the issues involved, develops a set of five criteria for assessing challenges of scaling, and then uses three programs widely discussed in the U.S. as examples of the challenges involved: Teach for America (an approach to teacher development), KIPP (a whole-school reform model) and the Harlem Children's Zone (a school-plus-community model). Five criteria are applied to assess scalability: cost, human capacity, tools and infrastructure, political support, and external or non-school factors.

Many innovations appear to have significant additional costs—for example the additional services offered by HCZ or the longer day and year in KIPP. The non-financial challenges, however, such as being able to find enough highly skilled people, can be just as significant and are often underestimated in discussions of scaling.

One cannot assume that a popular innovation necessarily represents a potential system-wide solution. Every school system should have a way to assess the potential value and challenge in adopting innovations. The goal should be to avoid either excessive enthusiasm or excessive skepticism, but to maintain a reasoned approach that over time yields collective learning. More independent research on the costs and benefits of major innovations is also required.

Recommendations

- All parties should avoid the temptation to proclaim small-scale innovations as the solutions to large-scale problems in education; rather they should be seen as promising ideas requiring further study before widespread adoption.

- There should be a more rigorous process for evaluating promising innovations to determine both costs and benefits, involving fair evaluations done by independent and neutral parties.
- There should be full and open access to data on costs and outcomes of innovations.

WHAT DOES IT TAKE TO SCALE UP INNOVATIONS? AN EXAMINATION OF TEACH FOR AMERICA, HARLEM CHILDREN'S ZONE, AND THE KNOWLEDGE IS POWER PROGRAM¹

Introduction: Why Discuss Scaling of Innovation?

This paper discusses the problem of scaling innovation in education in the United States. Why should one care about more innovation? After all, it could be argued that education has a long history of innovations being introduced with little or no lasting impact on the school system as a whole,² so that innovation is overrated as an improvement strategy. But the reality is that there are many calls³ for more innovation to improve education in the U.S., from the president down to local levels, asserting that it is only through more innovation that real improvement can occur. Most innovations start in one school, district or community. Although it is interesting and potentially important to see approaches that work well in a single setting, to make a difference at a system level—meaning a state or the nation as a whole—these ideas must be successfully adopted widely. This requirement draws attention to the problem of scaling.

The last 20 years have drawn increasing attention to this subject.⁴ This paper begins with a general discussion of the issues involved, develops a set of criteria for assessing scalability, and uses three programs that are currently widely discussed in the U.S. as examples of the challenges involved: Teach for America (an approach to teacher education and development); KIPP (a whole-school reform model); and the Harlem Children's Zone (a school-plus-community approach). Note that this brief takes no position on the effectiveness or value of these programs. Each is, in fact, the subject of considerable study and debate, with conflicting claims being made as to their results. They are used in this paper only to give concrete expression, through examples of different kinds of programs, to the challenges of scaling innovation.

It may seem odd to discuss scale without attention to which problems one is trying to solve, which practices to adopt, or to the effects of any given innovation. In reality, the choice of an innovation (or other policy) should depend much more on its purpose and demonstrated impact than on its scalability. As noted later, it is important to consider potential policies or innovations in light of good evidence on their effects. But because innovation is so widely advocated in education in the U.S. today, and because some commentators seem to suggest that the adoption of a few innovations could solve the problems of education in the country, it seems relevant to examine the challenges facing any innovation intended to have a national impact.

Scaling Innovation as Adoption and Adaptation

In a classic paper,⁵ James March distinguished between “exploration” and “exploitation” in the work of organizations. He argued that in order to grow and develop, organizations must explore new products or services or modes of delivery—in other words, innovate. In order to be achieve their goals, however, organizations must perform core functions well every day, and must ensure that what they *learn*—their exploration—becomes a central part of what they *do*—is exploited. While March does not specify a balance between these two functions, one could infer that he believes a fairly small amount of exploration must be levered into a much large amount of exploitation. That is the essence of scaling innovation.

Typically when people think about scaling innovation, they may think of models such as franchises or other ways of spreading successful products or approaches across an entire industry. It turns out, though, that scaling innovation is a problem in all fields, whether in the public or the private sector, and is complex in any field.⁶

The Decentralized U.S. Education System

To understand the implications of scaling up innovations, it’s important to understand the cost and structure of education in the United States. Let’s take cost first. According to the National Center for Education Statistics,⁷ the U.S. spent about \$610 billion on schooling in 2008-2009, of which about \$530 billion was for operating costs. This amounted to about \$10,500 per student. While per-pupil spending varies considerably among districts, a district with 10,000 students spending at the average level would have an annual operating budget of more than \$100 million. In other words, what seems like a large amount of money—even several billion dollars—is only a small proportion of the overall cost of schooling.

Structurally, scaling innovation in education in the United States is challenging because the system is large and decentralized. The United States has nearly 100,000 schools in some 14,000 districts in 50 states. Federal influence is mainly indirect, making it a very difficult proposition for the federal government to change practice in large numbers of schools. States remain the key actors, but while some state governments are powerful influences, in other states they are much less so. The limited capacity of states to support changes in school practice has often been noted.⁸

School districts, which employ the teachers and own the schools, vary in size from very large (more than a million students) to very small (a few hundred students); they also vary greatly in their ability to support lasting change in their schools. Schools additionally remain vulnerable to sudden or frequent shifts in direction based on changing political currents. A new superintendent or governor can mean a whole new set of policy directions, without enough consistency and stability for any of them to be well implemented.

Teaching as an occupation also has some features that make scaling of practices difficult. First, it is a mass occupation. Approximately 3.1 million teachers work in public education

in the U.S. Second, many districts and states have little common curriculum or teaching practice. Unlike other professions, teachers typically do not have common guidelines for practice except at the most general level (e.g. “know your subject”), although some efforts are being made in that direction through the National Board for Professional Teaching Standards, the Common Core, the National Council of Teachers of Mathematics standards, and others.

Fidelity Versus Flexibility

In short, changing practice in large numbers of schools in the U.S. is a very daunting proposition. The problem is actually more difficult, though, because scaling innovation in education also involves a fundamental tension between replicating a program or practice exactly and adapting it to meet different local circumstances.⁹ On the one hand, spreading an innovation necessarily means putting into practice key features of that innovation. If a new site does not use all, or virtually all those practices, then it can be argued that it is not

Scaling innovation in education in the United States is challenging because the system is large and decentralized.

really replicating the innovation. Local people may feel that a practice has been adopted when an outside observer would conclude that what is in place is not really the original idea at all. The cogent analysis of changes in teaching mathematics by Cohen and Hill¹⁰ is a classic example of teachers believing they were carrying out the new practices, when external observers concluded that they were only adopting some of its less important trappings and not the essence of the new approach.

This focus on fidelity remains a primary challenge in scaling up innovations. The New American Schools initiative put very substantial funding and effort behind a set of whole-school reforms, yet these reforms soon died away, leaving little in the way of results.¹¹ Writing about Success for All (SfA), one of the most widely adopted innovations in education, Slavin *et al.*¹² note the challenge of ensuring that the fundamental aspects of the program were actually put in place in all the sites that claimed to be doing it. In their discussion of the history of SfA, Peurach and Glazer¹³ point out the same challenge and show how difficult it was for the program to maintain a consistent approach as the number of sites using it grew rapidly.

On the other hand, according to another central theme in the literature on scaling, total replication may be an ineffective strategy. As a number of authors note, achieving the purposes of a program demands the thoughtful application of its principles rather than a mindless following of routines.¹⁴ Coburn’s frequently cited paper on scaling¹⁵ talks about breadth, depth, sustainability and shift in ownership to the local level as key indicators of scale. Consider, for example, the routine processing of service transactions, such as retail check-out or customer service. Good companies constantly train their staff in how to

respond to customer inquiries (for example go with someone to find an article rather than telling them where it is). They expect staff to carry out those behaviors regularly, but also to respond to the particular needs of each customer rather than insisting on an inappropriate behavior that is in the training manual. Yet even achieving consistency in these relatively simple operations is very difficult to do, as shown by Gawande's¹⁶ discussion of how difficult it is to have staff consistently wash their hands in hospitals.

In the case of education innovations, the requirements for adaptation are more stringent because the practices are much more complex and the settings in which they are applied can be highly variable.¹⁷ Virtually all significant innovations in education, including the three discussed in this brief, require a complex set of behaviors, not just from individuals but across an entire organization. Introducing a new approach to relating to students, for example, as all three of these innovations attempt to do, asks teachers to change their attitudes and behaviors, but may also require changes in the way the school organizes and operates itself, the way school leaders do their work, relationships with parents, work by support staff, and so on. Most of these changes cannot be specified in advance, either, but have to be worked out on the ground.

Another problem in scaling, outside the scope of this discussion, is that it is not possible to know if all parts of the innovation are equally important or powerful; if they are evaluated at all, the programs are typically evaluated as wholes rather than by assessing each component. Yet it may be possible to achieve most of the positive effects using simpler models or only some components. In any case, innovations in schooling are rarely reducible to a set of rote practices to be applied no matter what; they inevitably require thoughtful consideration about what is best to do under any given set of circumstances, which may vary depending on the students, skill levels of staff, the degree of community support, the demographics of the program participants and the community, and more.

Accordingly, scaling of innovations presents a dual challenge. On the one hand it is important to maintain the key components of an innovation, as these can easily be watered down or abandoned. On the other hand, it is also important to apply the innovation in a way that is appropriate to each different setting and group of people so that it achieves its purpose. This tension is one reason that it has been so difficult to obtain large-scale and reasonably consistent scaling of innovations, from Title 1 to New American Schools to many others.¹⁸

How Big Is “at Scale”?

A further question has to do with what amount of adoption constitutes “scaling” an innovation. While some innovations could potentially operate in every school or district in the United States, most, including those used as examples here, are not designed to be universal, but are targeted at a particular need or niche. The analysis in this paper assumes that the goal would be to have each of these (or similar) innovations operate in at least a significant minority of schools that fit their target purpose. The analysis that follows is based on these innovations being in place to meet approximately 25% of the potential

demand. That is an arbitrary figure but constitutes much wider use than any of them is currently achieving, yet without expecting anything like universal adoption. Wider adoption would increase costs and other challenges, though it could also, in some cases, produce economies of scale. Because education is a labor-intensive undertaking, though, economies of scale are generally small if they exist at all. The details of how the potential scale of each innovation was determined are in Appendix 2.

Requirements for Scaling Innovations

The literature on scaling of innovations lays out many of the challenges, but it does not clearly define a set of considerations involved in moving a practice from a few settings to many settings. The generic literature on implementing change may offer some insights,¹⁹ but although every innovation is by definition a change process, scaling of innovation is a particular kind of change effort.²⁰ This brief proposes five factors, drawn from both an inductive and a deductive analysis of the literature, as essential to thinking about the issues of scalability. For each of these five, the degree of challenge can be determined by comparing an innovation with what we might call the “standard” model of schooling (recognizing that the “standard” model also varies from place to place). The more an innovation differs from the standard model, the harder it will be to scale.

1. Cost. Each innovation has costs attached to it. Understanding scalability requires understanding the costs of an innovation compared with those of the standard model. How much more is required per student or per school for any particular innovation? These costs can be divided into two types: one-time or short-term start-up costs to put an innovation in place, and operating costs to sustain it from year to year. Usually the latter are much more significant than the former.

Costs can include all the standard categories of education spending. Typically staff salaries and benefits are the largest items, since they account for about 80% of education spending. Other costs could involve facilities, equipment, training, support services, administration, transportation, etc.

A higher cost for an innovation is not necessarily a bad thing; a more expensive model can be more cost-effective if it results in better outcomes. Moreover, there may be economies of scale to an innovation such that the cost per student or school could decrease, although this is not automatic. Still, an innovation that has significant additional costs, even for start-up, will be more difficult to finance initially and therefore harder to implement in a large number of settings.

The spending figures for the innovations in this paper are for all sources of funds, public and private. Each of these innovations is currently financed to a significant degree from private sources, but no distinction is made in this analysis about public and private sources of money; because the amounts are too large for private support, it is assumed that if any of the example approaches were to be scaled significantly, most or all of the money would have to come from public sources.

While cost seems in many ways the most straightforward criterion to assess, it has proved quite difficult to obtain complete and accurate cost data for these innovations, as discussed in Appendix 2. Accordingly, the paper relies on estimates. While these could be wrong by a significant degree, they do indicate the order of magnitude of the cost of scaling these or similar innovations.

2. Human capacity. Does the innovation demand a significantly higher level of skill or commitment than is found in the system now? Every program, including the “standard model” of schooling, assumes the involvement of people who are reasonably competent and committed, even though that criterion can be hard to achieve in practice. A scalability issue exists, then, only if the innovation requires something well beyond ordinary competence.²¹ For example, an innovation might require that all staff be highly skilled at particular complex teaching practices at a level well beyond what is currently present among most teachers. Or it might require more hours of work and a higher level of dedication—for example being available on evenings and weekends. Or an innovation may produce negative emotions in teachers that require attention.²²

Innovations may also ask for behavior that people cannot currently do, and that are very hard to specify or teach. For example, many programs call staff to hold students to higher expectations. While everyone understands this idea in principle, it is very difficult to tell or show someone how to actually raise one’s own expectations. School leaders and supervisors themselves may not know either how to do do or how to help others learn to do it, making it very hard to achieve effective implementation.

Human capital is the most important single resource schools deploy. If an innovation demands a much higher level of human capital than is currently available in most schools, expanding that innovation more widely will be problematic, independent of any additional costs involved, as skilled people are by definition in short supply. The exception would be an innovation that assumes that these additional skills can be developed by existing staff given appropriate training, in which case the issue is one of training requirements rather than skill supply.

The question of human capacity arises for different elements of the education system. For example, an innovation might require particularly skilled leaders (political and educational), or key support people such as trainers and consultants, or rank and file educators and other staff. Some innovations may also demand unusual levels of commitment from parents or students. The more complex an innovation, and the further it is away from current practice in most schools, the higher the human capital demands.

3. Tools and infrastructure. These issues are partly addressed in earlier points. Some innovations require specific supports that are not normally available in schools. These could include particular kinds of facilities, technology, training materials, extra time for training and development, etc. These are typically not major items in terms of cost but can be difficult to find or provide in quantity. For example, an innovation that requires considerable internet bandwidth could be a challenge in schools lacking that capacity.

4. Political support. The political dimension is sadly neglected in many discussions of education reform. Reforms cannot be adopted, or sustained once they are adopted, unless there is ongoing support from elected leaders, from school and district leaders and, in the end, from rank-and-file teachers, students, and parents. It is possible to implement changes without that support, but not to sustain them or for them to be effective, as has been shown by a great deal of research and experience around the world.²³ It may seem unnecessarily time-consuming to engage these various groups as a way of building their support, but without that investment of time and energy, longer-term results will inevitably be disappointing.

External political support depends on people's perceptions of a particular innovation. Those perceptions are not always well-informed, but they are still real. Despite all the rhetoric of innovation, there is still a strong public belief in schools as they have been known or experienced by most people in the past.²⁴ So while very different models of schooling are frequently advocated, they rarely garner high levels of support from teachers

One cannot make any assumptions about whether particular innovations will be easier to implement at scale.

or parents. There is a conservative bias in public views of what kinds of schooling are appropriate, and a strong resistance by parents to the idea that their children will be the subjects of experimental approaches—even when it is clear that the current approaches are not working particularly well. Metz²⁵ referred to this issue years ago as a desire for “real school,” but the same phenomenon has been observed in political processes for centuries. We can also see this bias at work in the continued popularity of practices such as ability grouping or retention in grade, well after both of these have been shown to be ineffective by considerable evidence. Innovations that do not meet the public-acceptability test are unlikely to succeed at scale.

5. Non-school factors. Factors outside the school can affect the scaling of any initiative. For example, Lee and Luykx²⁶ show how implementation of a science program in multiple sites is affected by the demographic characteristics of the students. An intervention that works for African-American students might not be equally successful with Hispanic students, and so on. Or the success of a particular innovation might require considerable stability in the student body and so would not be effective in areas of high mobility. Or it could require considerable parental time and commitment that are simply not available in some communities, or broadband access at home for families that do not have it and cannot afford it, or it might presume a high degree of collaboration between schools and social services when such collaboration is hard to develop and sustain.

Finally, it cannot be assumed that the difficulty of any of these challenges will change in lockstep with an increase in scale of application. Some issues might be easier with wider implementation. Political acceptability may increase, or programs may benefit from economies of scale. In some cases, as innovations become more mainstream they could replace other current activities, thus reducing their additional cost. However in other areas

the challenges could increase with greater scale. For example it is harder to maintain a particular organizational culture when there are many more sites involved. Or if an innovation requires staff with particular attributes, it may be more difficult to find those people in much larger numbers. One cannot make any assumptions about whether particular innovations will be easier to implement at scale.

Applying the Criteria to Real Examples

Abstract discussions of scalability are useful, but as with any such discussion, the application of the principles to real circumstances is required in order to understand the phenomenon fully. For this purpose, we now apply these criteria to three innovations currently much discussed in the United States: Harlem Children's Zone (HCZ), KIPP (Knowledge is Power Program) and Teach for America (TfA). These three were chosen because they are prominent but also quite disparate kinds of innovations. KIPP is a model of school delivery; HCZ is a community-based approach, and TfA does not deal with school delivery but with training teachers and developing leaders. So these examples should raise different kinds of issues about scalability.

Many other innovations might equally have been chosen. In choosing these three we are not endorsing them or suggesting that they are better or more significant than other reforms, nor are we critiquing or evaluating them in any way. This report makes no inferences about the value, cost-effectiveness, desirability or any other attribute of these programs. There is an active debate about the value of each program, with an increasing number of studies attempting to assess their impact on a range of outcomes. They are included here only as prominent current examples to illustrate issues of scalability for innovations in education.

The five scalability criteria are described and applied to the three innovations in Table 1. A fuller description of each innovation is in Appendix 1, while Appendix 2 gives more details on how these estimates and judgments were made.

Data for the analysis was drawn from public domain sources, including the reports and websites of the three innovations. None of the organizations provides full information on their websites regarding costs or other scalability requirements, such as human capacity issues or required political support. We contacted each of the three sponsoring organizations and invited them to provide additional information in relation to these points, but none of them did so. Readers should be aware that each of these innovations is quite complex, involving many elements that are not fully described here. Those wanting a deeper understanding of any of these innovations should look at other sources, some of which are referenced in Appendix 1.

In this paper conservative estimates are used in all cases; it is likely that the actual scaling requirements would be greater than are described here.

Table 1 : Summary of scalability requirements for each innovation

Innovation	Integrated services	Intensive charter school	Alternative teacher education and leadership development
<i>Example program</i>	Harlem Children’s Zone	KIPP	Teach for America
<i>Key program features</i>	Provide children, families in defined geographic area with comprehensive education and social services, preschool through adult; may also involve charter schools	Charter school model focused on children in poverty; longer school day and week; strong focus on college preparation and strength of character	Recruits, trains high-achieving college students to teach 2 or more years in high-need schools with desired ongoing leadership role in education; highly selective; summer training plus ongoing development
<i>Current scope</i>	Harlem, New York; approx. 10,000 children/youth	125 schools in 20 states with 40,000 students	5000 trainees/year; much larger number of alumni
<i>Potential scope</i>	100 zones could serve about 25% of target population	25% of “high need” schools would be about 5,000 schools in US	35,000 new teachers/year would meet 30-40% of new teacher demand in high need schools
<i>Costs</i>	Estimate model requires about \$5,000 (50%) more per student per year primarily for additional non-school services; overall: \$50m/zone/year, or \$5 billion for 100 zones	Estimate model requires about \$2,500 more per student per year, primarily for extended day/year costs; for 5,000 schools, 1.5m students, about \$4 billion	Estimate 35,000 teachers/year would require about \$600million; some offset from reducing traditional programs; higher turnover could bring additional costs
<i>Human capacity</i>	No unusual requirements except perhaps exceptional CEO (probably not a scale issue)	Requires exceptionally skilled, committed teachers; could be hard to find 100,000 even with additional pay	Likely not enough candidates with desired qualifications to reach target number above.
<i>Tools and infrastructure</i>	Hard to create and sustain systems linking schools effectively with social services	Systemic effort required to create new forms of pedagogy and teacher-student relationships	No unusual requirements.
<i>Political support</i>	No unusual requirements	No unusual requirements if implemented widely	Highly controversial approach would be problematic to scale
<i>External requirements</i>	Creating and sustaining appropriate extended services is difficult	High level of desired parental commitment a challenge at scale	None other than those already noted
<i>Overall scalability analysis</i>	Biggest challenge: additional cost; second challenge is making integrated services work effectively	Biggest challenges: additional cost, finding enough appropriate staff and families	Biggest challenges: finding enough qualified applicants, retaining them in schools; financial requirements hard to determine
<i>Other comments</i>	Each community would be different, so economies of scale likely small	Challenges of quality and consistency get larger as number of schools increases	Key question: Is this an add-on or a fundamental change in teacher preparation

Finally, this discussion does not address the issue of fidelity. Although that remains a central question in the broader discussion of scaling of innovation, the assumption in what follows is that the essential elements of innovations would be scaled in ways that preserved the innovations' central character and, in particular, their efficacy.

Conclusions

The analysis in this paper is indicative of the requirements needed to scale various innovations to a level at which they would have a significant national impact. The actual requirements for any particular innovation, including the three discussed above, could vary widely from the numbers in this paper, but the orders of magnitude are probably approximately right.

There appear to be significant challenges in moving any innovation to a level that would reach even a large minority of the students across the country for whom the program is intended. The framework of five factors introduced here (cost, human capacity, specialized tools, political acceptability and non-school factors) seems a useful way of thinking about these challenges. It should be developed more fully and applied more widely by other researchers and policy analysts.

Second, many innovations do appear to have significant additional cost requirements that would affect their scalability. All three of the examples here do require additional resources. The scale of additional resources can be large in absolute terms (e.g. billions of dollars) while still being small in terms of the overall size of the education sector in the United States (more than \$500 billion). More expensive programs can also be more cost-effective depending on their results. It would be useful to conduct more detailed cost-benefit studies of these or other innovations. Interestingly, in all the discussions of these programs, few include any analysis of costs, and even fewer attempt to assess costs against benefits.

Third, the non-financial challenges of scaling can be as significant as the cost issues, or even more so, yet they are often underestimated in discussions of scaling. Some innovations depend on unusual levels of staff skill and commitment, which are very hard to replicate on a much larger scale, as is organization culture generally. Some face significant political controversy and challenges in wider adoption, while still others depend on requirements, such as high levels of interagency collaboration, that are hard to attain and to sustain. Innovations would benefit from being more explicit about what it is that they require that differs from the "standard model" and what is involved in producing and sustaining that difference.

Recommendations

One key implication of this brief is that one cannot assume that an existing innovation, no matter how promising it may be in a few settings, necessarily represents a potential

system-wide approach or solution. In general approaches to innovation in education lack discipline and rigor; the system as a whole does not have a well-organized approach to research and development,²⁷ yet one is badly needed. While it is important to continue to experiment with interesting ideas, that experimentation needs to be accompanied by independent, high-quality research and should always have at least one eye towards the potential wider application of the innovation. Governments, foundations and others could collaborate more effectively to ensure that disciplined innovation produced more fruitful learning.

It is also important that the organizations developing innovations make their data available to researchers. Proponents may be reluctant to do so if feel they are being attacked while working hard for something in which they believe. Yet without open data it is impossible to assess the value and results of these programs. Proponents of innovations should also be encouraged to think about scaling issues and to take a position on the scalability of their innovations. Do they see their model as potentially applicable in all appropriate settings, or is it designed to be used in only a few places?

Policy-makers should be cautious about adopting innovations no matter how widely promoted they are. Education has a long history of adopting fads without much supporting evidence and abandoning them soon after,²⁸ an approach that wastes resources and corrodes professional confidence. Every school system should have a way to assess the potential value and also the potential cost or challenges involved in adopting major innovations; an analysis such as that in this paper could be a starting point. The goal should be to avoid either excessive enthusiasm or excessive scepticism, but to adopt a reasoned approach that over time will yield collective learning.

A further implication is that public debate on improving education is not well served by advancing isolated cases as system-wide solutions. The fact that something good can happen in 10 or 50 schools does not mean that the same thing can be made to happen in 5000 schools. It would be helpful if policy analysts were less inclined to use such small examples as the basis for claims about what is possible at a national level.

Finally, disciplined innovation calls for different responses on the part of the research community. Too much of the discussion and analysis of innovations, including these three, appears to lack objectivity and to start from a position either for or against the particular innovation. While it is impossible to remove points of view from research, the goal should always be dispassionate assessment of the evidence. An organization like the American Educational Research Association could, for example, usefully develop protocols for the fair evaluation of innovations. The foundations and other private sources that often provide initial support for innovations, while understandably wanting to see their “babies” succeed, need to adhere to fair and independent evaluations with open access to data.

Innovation is necessary to learning and improvement, but only if it is undertaken in a thoughtful, open-minded and evidence-informed way. The goal of this brief is to make a contribution to that discussion.

- All parties should avoid the temptation to proclaim small scale innovations as the solutions to large scale problems in education; rather they should be seen as promising ideas requiring further study before widespread adoption.
- There should be a more rigorous process for evaluating promising innovations to determine both costs and benefits, involving fair evaluations done by independent and neutral parties.
- There should be full and open access to data on costs and outcomes of innovations.

Appendix 1: Descriptions of the Three Programs

Harlem Children's Zone (HCZ)

Harlem Children's Zone is a non-profit organization that provides a wide range of social, communal and educational services to families and children within a 100-block area of Harlem, New York. The organization aims to address needs around housing, crime, health and other critical issues affecting community members. Program offerings include several that follow children from early childhood to high school graduation (referred to as the "HCZ pipeline"), such as Harlem Gems, an all-day pre-k program, and other services that "reweave the social fabric" of the community in health, culture, and education, including a fitness program that offers free classes in karate, fitness, and dance, an employment and technology center to teach computer and job-related skills, and Boys to Men and Girls to Power Leadership programming, amongst many others.

HCZ also operates its own charter schools, though most of the children it serves attend regular public schools. The HCZ Promise Academies (charter schools) have an extended school day and year, with additional out-of-school supports such as tutoring services and courses in mathematics and literacy on Saturdays.²⁹ The program is free to children and families.

HCZ programming is based on a belief that the provision of community services can have strong influences on student achievement. Therefore, surrounding young people with enriching experiences and supports will lead to better social and educational outcomes. When asked about the organization's vision of success, CEO Geoffrey Canada has stated that "having a young person graduate from college" is the ultimate indicator of success for the organization. Several interim indicators may also suggest whether the organization is on target (e.g. test scores, students performing at grade level). Studies have sought to explore the relationship between communal supports and educational outcomes using HCZ as a case study, including Dobbie and Fryer,³⁰ who argued that high-quality schooling alone can close achievement gaps and reduce the effects of poverty, and a Brookings Institute report³¹ that questioned the value of community service investments.

In 2010, the Obama administration accepted the first round of applications for 20 grants in its Promise Neighborhoods program, whose goal is to replicate HCZ's success in other impoverished areas. Several other organizations in cities around the country and the world have put in place community-based service models similar to HCZ, including Boston, Los Angeles, New Orleans, Israel, and South Africa.³²

KIPP (Knowledge is Power Program)

KIPP operates a national network of free, open-enrollment charter schools with a focus on college preparation and developing character strengths. Founded in 1994, KIPP currently has 125 charter schools in 20 states and the District of Columbia, with about 40,000

students. KIPP schools are structured around an extended school day (Mon-Fri 7:30am-5pm and Saturdays from 8:30 am to 1:30pm). The schools follow the same state and federal curriculum standards as other public schools. There are a range of class sizes amongst KIPP schools, but the average is 16.³³

Any student can enroll in a KIPP school who meets the district's residency requirements; acceptance is dependent on lottery selection. Admissions are open regardless of the student's prior academic background, conduct or socio-economic status. Parents of KIPP students must sign a contract agreeing to certain KIPP school procedures, which include expectations for time scheduling, dress code, and so on. KIPP says that 87% of students at KIPP schools are low income, 85% qualify for federal free or reduced meals, and 95% are African American or Latino.

Across all KIPP Schools, about 38% of teachers are African-American or Latino, 32% are Teach For America alumni, and 32% hold master's degrees. As of August 2011, 73% of KIPP school leaders were former KIPP teachers, and 65% held master's degrees. Hiring decisions are made locally; teachers are usually required to have a minimum of two years of teaching experience in a low-income school, but KIPP also offers specific development programs for new teachers who do not yet have classroom experience.³⁴ KIPP also actively works to develop teachers to lead and start their own KIPP schools.

Teach for America (TFA)

Founded in 1989, Teach for America is an independent non-profit organization that develops, trains and places recent college graduates as teachers in low-income, high-need areas of the United States with the goal of developing its members into education leaders in schools and communities. The program has a strong belief that students can achieve at high levels despite poverty, and that high-achieving and passionate teachers can help these students succeed. In its first year the program placed 500 teachers in 6 regions of the country; by 2012 that had increased to 5800 teachers in 46 regions.³⁵ In 2007, founder Wendy Kopp also started Teach for All, an organization that seeks to spread the TFA approach to other countries.

The program is competitive and recruits from recent university graduates. Applicants must have earned a minimum 2.5 GPA and be a U.S. citizen or have permanent residency status. In 2012 the program received 48,000 applications for 5,500 spots (around an 11.5% acceptance rate). 17% of recruits work full time prior to joining³⁶ According to its website, 2011 incoming corps members were 77% graduating seniors, 13% professionals; 34% "people of color." According to the TFA annual report, more than 7,000 of 24,000 alumni are still teaching while over 660 serve as school or system leaders.³⁷

New members participate in a five-week intensive summer training program in one of nine locations; during training, the recruits prepare for their first year of teaching and get acquainted with the region in which they will be working. Corps members teach summer

school for four of the five weeks of training and receive coaching. Training includes observations, rehearsals and reflection, lesson-planning clinics, curriculum sessions.

Teachers are paid as full-time employed teachers in their districts. Compensation varies by region and ranges from \$25,500 to \$51,000, including benefits. Corps members are considered to have “alternative” certification and depending on the region must continue coursework towards their next level of certification.

Appendix 2: Details on Applying the Criteria

Children's Zone

Scope for scaling

Harlem Children's Zone serves about 10,000 young people in a defined high-poverty area of New York City. In 2007 there were about 13 million children in the United States living in poverty, with nearly half of them (approximately 6 million) in "deep" poverty (below 50% of the poverty line).³⁸ About 80% of Americans live in or near urban areas, so there would likely be 4-5 million very poor children living in these areas. To have a national impact it seems reasonable to say that an HCZ model should reach at least 25% of the poor children in urban settings, or about 1 million children. Based on about 10,000 children per "zone," and assuming (which is doubtful) that 100% of participating children in any "zone" would be living in deep poverty, this means that a minimum requirement would be approximately 100 "zones." These zones would presumably need to exist in every major urban area in the United States, with several zones in the largest cities.

Application of criteria

1. Cost. HCZ does not provide cost data on its program on its website, nor do any of the public analyses or evaluations of HCZ provide data on the cost of delivering the program. Costs are very difficult to estimate for this model because it has many components, in schools and beyond, some of which use services already available in the community and others of which use new services. The additional costs of providing a children's zone would include:

- the operations of the central organization itself, including operation of the board, financial administration, communications, fundraising, coordination, HR and evaluation;
- any additional costs for existing programs, such as a longer school day or year in its charter schools;
- the operation of services that would otherwise not be available in these communities, including most if not all of the additional social services, from early-childhood provision through youth leadership and parent engagement.

HCZ does report³⁹ that it serves "more than 10,000 children and more than 7,400 adults.": The FY 2010 budget for the agency overall is over \$75 million." Dividing the \$75 million by 10,000 children yields a per-child cost of \$7,500 per year. This budget supports a rich

array of programs and a large central organization; it might well be possible to provide a children's zone with many of the same programs for significantly less money. Moreover, there would likely be some economies of scale if similar programs operated in many places. For purposes of this analysis we will use the figure of \$5,000 per target student, or approximately 50% above current average per-pupil spending levels for schools, most of which is presumably for social services, not schooling per se. Nevertheless, the 2008 auditors' report on the Promise Academy, an HCZ charter school, indicated that the school's regular funding of approximately \$6 million was supplemented by about \$3 million in additional services provided by HCZ: close to 50% beyond regular funding.

Assuming 1 million students as the scale target, implementation of about 100 children's zone programs across the United States could reasonably be expected to require \$5 billion per year in additional costs, most of which would be used to provide extended services such as early childhood programs. Whether this amount is seen as "large" or not is a matter of judgment. While \$5,000 per student is a very large increase over current average spending on education in the U.S., \$5 billion is only about 1% of total operating expenditure for U.S. public schools.

2. Human capacity. The HCZ model does not seem to have any special human capacity requirements. HCZ has a particularly charismatic leader in Geoffrey Canada, but if there were 100 such zones they would only require leaders who were reasonably effective in leading and managing local programs. Similarly, the model appears to assume that staff in the various programs require only a reasonable level of commitment and skill that any similar program would expect.

3. Tools and infrastructure. The main requirement that is unusual in the HCZ model is close working connections between schools and external agencies and organizations. Creating and maintaining such connections, while often recommended, has proved to be extraordinarily difficult. Coordination of services also carries a cost in terms of people's time. Many previous efforts to create similar initiatives in other places (for example "education priority zones" in France or "education action zones" in England or previous attempts in the U.S.) have proved unable to sustain themselves or to demonstrate success.⁴⁰ Building these relationships and keeping them effective does seem to require a significant additional "infrastructure" in terms of meetings, teams, political capital and relationships.

4. Political support. The ideas behind HCZ would only be controversial in communities with highly conservative values, where some of its emphasis on empowering youth, working with single mothers, etc. might be seen as threatening. However these ideas are quite consistent with mainstream American thinking and are popular in the communities involved, and HCZ itself has been very successful in garnering public and community support for its work.

5. External requirements. The HCZ model assumes that a community either has or can build a strong network of additional supports for young people and families.

Overall: The largest barriers to scaling HCZ to 100 communities would be the additional funds required, and the task of building and sustaining effective program connections between schools, families and social services.

Intense charter school

Scope for scaling

KIPP seeks to prepare students in underserved communities for success in college and in life, with an emphasis on partnership among parents, students, and teachers that focus on learning, outstanding educators, more time in school, and a strong culture of achievement.

The potential audience for a program like KIPP is larger than the potential audience for a children's zone because the latter requires a high concentration of students in an urban areas, while a KIPP school could operate in virtually any community with a significant concentration of needy students. The average KIPP school has about 300 students (though many will get larger as they add grades moving towards full implementation), but schools operating the program could presumably be larger or smaller.

There are nearly 100,000 schools in the United States. Assuming that the intention was to have this kind of charter school model operate in 25% of high-need schools across the country, that would be roughly 5,000 schools, or 40 times as many as KIPP presently has. This would serve about 1.5 million students and require approximately 100,000 teachers. Such growth in a single model of schooling has not happened in recent history. Success for All, the school model that displayed the most growth in the United States in recent years, grew to a maximum of 1,600 schools over a 15-year period⁴¹ and then dropped to about 1,200.

Application of criteria

1. Cost. There is a lively debate about whether KIPP schools spend more than similar schools. This is a complex question to resolve for several reasons. KIPP itself does not publish figures on total per-pupil spending in its schools because, they say, these figures vary greatly from state to state and do not include strictly comparable elements. Rather than try to determine a cost per student for KIPP, I take a different approach. KIPP operates a longer school year—as much as 600 extra hours per year for students,⁴² something like 50% more time than the typical U.S. school's 190 days of 5.5 hours a day. One report finds that KIPP teachers average 65 hours of work per week.⁴³ Whether KIPP is currently able to do this for around the same cost as regular schools is immaterial to the scaling challenge, since it seems inconceivable that operating this model in thousands of schools could be done without hiring more staff or paying staff more for a much longer work week. A reasonable estimate would be that the cost of this increase would be about

\$2,500 more per student per year, or 25% above current total per-pupil spending (assuming a cost of \$10,000 per student per year, of which about 60% is teacher salaries, which would then need to be increased by about half to pay for the additional hours, discounted somewhat because not all teachers would need longer work weeks). Applied to 1.5 million students, this translates into an increase of nearly \$4 billion. This does not include any additional funds for any purpose other than more staff time, however, such as the additional support components for schools and students required by the KIPP model. The KIPP Foundation, spent more than \$36 million in 2011 supporting the organization and its schools, either directly or indirectly,⁴⁴ or about \$250,000 per school. If there were 5,000 such schools, the per-school cost would probably be lower, perhaps \$200,000 per school for 5,000 schools, which adds another \$1 billion to the national cost. So operating a program of this kind at this scale might require an additional \$5 billion nationally.

While finding or renovating decent school buildings has been a major cost issue for KIPP, this analysis assumes that these additional schools would be housed in existing buildings, so no additional capital costs are included.

2. Human capacity. KIPP prides itself on recruiting teachers with very high levels of commitment to children. A high proportion of its teachers are, in fact, Teach for America graduates. At the same time, KIPP asks its teachers to work significantly harder—or at least more—than is the case in most schools. It is widely acknowledged that teaching in schools with many high-need students does pose additional demands on teachers and makes their work more difficult (if also potentially more rewarding). There is a debate about whether KIPP staff currently leave teaching sooner than other teachers, but in any case it is hard to imagine that this level of effort and dedication could be sustainable across 5,000 schools employing maybe 100,000 teachers. There must be a real question about whether it would be possible to find that number of teachers willing to make the commitment called for in this model, even if their pay were increased commensurately with the additional hours. Nor is it certain that higher pay would succeed in attracting teachers motivated primarily by the challenge of the work, as KIPP wants.

A model that depends on an especially high level of commitment and skill by staff and leaders must face scale challenges about whether enough people of the desired caliber can be found and retained.

3. Tools and infrastructure. A model, such as KIPP, with a particular approach to teaching, learning, and student-teacher relationships that is not standard in most schools, will necessarily require additional training for staff and new kinds of systems to support these practices. There is always a challenge in moving a large number of organizations to a new standard of practice; extending those characteristics across 5,000 schools would be especially challenging. While it requires money, even more it requires building the appropriate culture, having the appropriate training, keeping the right leadership focus, and aligning other systems such as pay and evaluation, so that all these factors support the central goals. This is not an area where there are economies of scale, either; the demands for training and support will likely increase more than proportionately as more schools are added, as there will inevitably be some schools that run into serious difficulties.

4. Political support. Political support for a model similar to KIPP does not seem to be an area of difficulty. Opposition to KIPP and similar charter school models is typically centered on their effect on the public system or on financial and admission issues, and these would be very different if there were 5,000 such schools. Nobody objects to a model with high expectations and strong teacher-student connections, nor to the additional hours and services, if they are not imposed on staff. The absence of elements that are likely to be unpopular suggests there do not seem to be reasons to expect difficulties on this dimension.

5. External requirements. KIPP requires a considerable degree of commitment on the part of students and parents. Indeed, an ongoing criticism of the program has been allegations about selectivity in admissions and high levels of attrition. Whatever the merits of those criticisms, the KIPP model does have requirements for students and their families that would not be met by all those in its target group.

Overall: The main barriers to scaling a program like KIPP would be the additional cost and the difficulty in finding both enough staff and enough students or families with the requisite commitment to make the program successful.

Alternative teacher/leader development

Scope for scaling

There are 3.1 million teachers working in the United States' public schools. Teachers are one of the largest occupational groups in the labor force. The U.S. Department of Education reports teacher turnover in the nation's public schools in 2008 at 8%,⁴⁵ yielding an annual demand for 250,000 new teachers per year. Another source⁴⁶ estimates that urban and poor communities in the U S requires about 700,000 new teachers per decade, or about 70,000 per year. That figure is about 30% of the total turnover figure, so it seems consistent. As noted below, however, it seems likely that turnover would be higher among TFA teachers, requiring more to be trained to fill the same number of positions on an ongoing basis. If Teach for America were to meet 30-40% of the demand for new teachers in high-need schools, it would have to train approximately 35,000 teachers per year — seven times its current output, or about 14% of all new teachers in the country. This analysis also assumes a perfect match between supply and demand; in most cases a labor supply system will produce somewhat more supply than is needed in order to offer employers some choice.

Application of criteria

1. Cost. An alternative certification program such as TFA could require extra spending in three ways. First, there is the actual cost of the training. How much does it cost to prepare

a teacher through TFA compared with other routes? Second, school districts pay a fee per teacher when they enter into a contract with TFA, an additional cost that does not occur with traditional certification programs. Third, if alternative certification graduates leave teaching at significantly higher rates than do other teachers, that imposes a cost related to the program. Of course if alternative certification graduates are more effective teachers, that could compensate for the additional costs.

The evidence is not clear on any of these points.

In terms of cost of training, a key fact is that in an alternative certification program such as TFA, all participants already possess a four-year college degree. Since many if not most teachers obtain regular certification as part of a four-year degree, a post-degree model, even if only a few weeks long, will tend to be an additional cost to the teacher-preparation system as a whole. According to TFA's 2011 audited statements, the organization's budget of \$220 million was in several major categories: recruitment and selection (\$38 million), pre-service institute (\$33 million); placement and professional development (\$92 million); alumni (\$21 million), and management and fundraising (\$36 million). As one example of an additional cost, according to Goetz & Aportela,⁴⁷ TFA employs 100 recruiters to search out and work with potential candidates. A reasonable estimate would be that the cost of initial training for 5,000 new teachers would include the first two categories and part (let's say 25%) of the placement and professional development figure—for a total of more than \$90 million, or about \$18,000 per new teacher. This puts training at only 40% of total TFA spending, which seems conservative. Training seven times as many teachers would require about \$600 million per year. Some of this could be offset by reductions in traditional teacher-training programs that would no longer be required. It is impossible to estimate these savings since the costs of teacher education vary greatly by institution; if most of the consequent reductions came in low-cost institutions, the offset would be smaller. Further, a substantial portion of students in traditional teacher-education programs do not end up teaching at all, which creates additional costs; if a higher percentage of TFA graduates, in comparison, do end up teaching, that would generate some efficiencies. A reasonable estimate of net new costs for a much larger TFA approach, given all of the above, might be about \$300 million.

TFA charges the school districts it contracts to supply teachers to a fee of approximately \$5,000 per teacher. Some part of that may offset some district costs for professional development, which TFA provides to all its alumni on an ongoing basis. If such a model provided 35,000 teachers a year, the cost to school districts would be about \$175 million; this money, however, would presumably offset some or all of the training costs already included above.

The third area has to do with attrition. TFA requires a two-year teaching commitment from its members, although many do continue to teach for longer or stay involved in education in other ways. Exactly how many stay and for how long is not clear from the available studies. The average teacher in the U.S. has between 10 and 14 years of experience.⁴⁸ TFA claims that in the high-poverty districts it currently serves, its graduates' average tenure is as long or longer than district averages. Whether or not this is true, if implemented at the

much larger scale anticipated here, the average tenure of TfA graduates would have to reach levels similar to the current U.S. averages. Under these conditions it seems virtually certain that the TfA model will result in higher attrition rates, thus increasing the number and cost of new teachers needed. A plausible estimate would be that if TfA provided 14% of all teachers, and the average tenure of TfA graduates was 5 years, total demand for new teachers would increase by about 10,000 per year. Recent studies⁴⁹ have estimated the total cost of replacing a teacher (training, recruitment, induction, etc.) at approximately \$40,000, so this higher turnover would add about \$400 million to the nation's education bill.

An estimate of the total cost of operating a model like TfA for 35,000 teachers per year, then, is about \$1 billion annually above current spending. A substantial portion of this could be offset by reductions in traditional teacher training, however.

2. Human capacity. A central purpose of TfA is to raise the quality of people going into teaching by recruiting very talented students who would not normally have considered this career choice. By definition, then, this approach requires a higher level of human capital than does the traditional system. TfA has tripled its enrolment since 2003 while apparently being able to maintain its standards in terms of the quality of applicants. However it is not clear whether enrolment can be expanded seven-fold while maintaining the same caliber of participants. As noted earlier, teaching is a mass occupation, accounting for about 15% of all college graduates, based on 1.6 million bachelors graduates per year⁵⁰ and 250,000 new teachers required per year. Bringing 35,000 of the top 10% of college graduates (160,000 people) into alternative certification would mean that more than 20% of elite graduates would go into teaching via TfA, compared with 3% currently. This seems highly problematic as a possible target. If TfA or an equivalent program had to be much more flexible in its admissions policy, the whole purpose of the program could be threatened.

3. Tools and infrastructure. An alternative certification model such as TfA does entail particular kinds of training and support for its members, but none of these appear to be unusual or suggest any particular constraints on its wider application. As in any program, expanding the staff of skilled trainers five- or six-fold would not be easy, but the skill set required, involving expertise in pedagogy and other aspects of teaching, would likely be reasonably available, including from TfA alumni.

4. Political support. TfA has been highly controversial since its inception, and the debate about it continues to be heated. TfA is strongly opposed by many teachers and teacher organizations because it appears to take the view that successful teaching does not require extensive training. Expanding this model dramatically would likely generate significant political opposition. Because success requires graduates of the program to be hired by schools and districts, rapid expansion could be politically problematic.

5. External requirements. The only real external requirement for TfA is the willingness of top college graduates to enroll in it and to take up teaching, at least temporarily, as an occupation. This has already been discussed under point 2.

Overall: The main barriers to scaling TfA are the availability of sufficient numbers of high-quality applicants and political acceptability. The size of the financial costs is difficult to estimate and could range from significant to quite small.

Notes and References

1 Research assistance for this brief was provided by Nathalie Carrier, a graduate student at the Ontario Institute for Studies in Education at the University of Toronto.

2 Cuban, L. (2008). The perennial reform. *Phi Delta Kappa*, 90(4), 240.

3 For example Hess, R. (2010). *The same things over and over*. Cambridge: Harvard Education Press.

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

Datnow, A. (2002). Can we transplant educational reform, and does it last? *Journal of Educational Change*, 3(3-4), 215-239;

Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10(2), 101-113;

Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, 13(2), 155-190;

Zavadsky, H. (2009). *Bringing school reform to scale: Five award-winning urban districts. educational innovations*. Cambridge, MA: Harvard Education Press.

5 March, J. (1991). Exploration and exploitation in organization learning. *Organization Science*, 2, 71-87.

6 Bradach, J. (2010). Scaling impact. *Stanford Social Innovation Review*, 8(3), 27.

Farrell, D., & Kalil, T. (2010). United States: A strategy for innovation. *Issues in Science and Technology*, 26(3), 41-50;

OECD (2005). *Governance of innovation systems*. Paris: Organisation for Economic Co-operation and Development;

Samoff, J., Sabatane, E. & Dembele, M. (2011). Scaling up by focusing down: Creating space and capacity to extend education reform in Africa. In L. Tikly and A. M. Barrett (Eds.), *Education quality and social justice in the South: Challenges for policy, practice and research*. London: Routledge.

7 National Center for Education Statistics (2012). Fast facts – Expenditures; summary and table reproduced from U.S. Department of Education, National Center for Education Statistics. (2012). *Digest of Education Statistics, 2011* (NCES 2012-001), Table 191 and Chapter 2. Retrieved March 8, 2013, from <http://nces.ed.gov/fasTfActs/display.asp?id=66>.

8 Le Floch, K., Boyle, S. & Therriault, S. (2008). *Help wanted: State capacity for school improvement*. Washington: American Institutes for Research. Retrieved March 8, 2013, from http://www.air.org/files/Research_Brief_I-State_Capacity_for_School_Improvement_091508_r1.pdf;

Murphy, P. & Oujidani, M. (2011). *State capacity for school improvement: A first look at agency resources*. Seattle: University of Washington, Center on Reinventing Public Education. Retrieved March 9, 2013, from <http://www.crpe.org/publications/state-capacity-school-improvement-first-look-agency-resources>;

Gottfried, M., Stecher, B., Hoover, M & Cross, A., (2011). *Federal and State Roles and Capacity for Improving Schools*. Santa Monica, CA: Rand.

9 Datnow, A. (2002). Can we transplant educational reform, and does it last? *Journal of Educational Change*, 3(3-4), 215-239;

Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, 13(2), 155-190.

10 Cohen, D. K. & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294-343.

11 Bodilly, S. J., (1998). *Lessons from New American Schools Development Corporation's demonstration phase*. Santa Monica, CA: Rand.

12 Slavin, R. E., Madden, N. A., Karweit, N. I., Dolan, L., & Wasik, B. A. (1992). *Success for all: A relentless approach to prevention and early intervention in elementary schools*. Arlington, VA: Educational Research Service.

13 Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, 13(2), 155-190.

14 Winter, S. G., & Szulanski, G. (2001). Replication as strategy. *Organization Science*, 12(6), 730-743.

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

16 Gawande, A. 2007. *Better*. London: Picador.

17 Datnow, A. (2002). Can we transplant educational reform, and does it last? *Journal of Educational Change*, 3(3-4), 215-239.

18 E.g. Borman, G.D. (2005). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. In L. Parker (Ed.), *Review of Research in Education*, 29 (1-28). Washington, DC: American Educational Research Association.

19 Fullan, M. (2001). *The new meaning of educational change*. New York: Teachers College Press.

20 Samoff, J., Sabatane, E. & Dembele, M. (2011). Scaling up by focusing down: Creating space and capacity to extend education reform in Africa. In L. Tikly and A. M. Barrett (Eds.), *Education quality and social justice in the South: Challenges for policy, practice and research*. London: Routledge.

21 Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Education Review*, 66(1), 1-26.

22 Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, 32(3), 303-319.

23 Mirel, J. (2002). Unrequited promise. *Education Next*, 2 (2), 64-72. Retrieved March 13, 2013, from <http://educationnext.org/journal/summer02/>;

Levin, B. (2008). *How to change 5000 schools*. Cambridge, MA: Harvard Education Press.

- 24 Mirel, J. (2002). Unrequited promise. *Education Next*, 2 (2), 64-72. Retrieved March 13, 2013, from <http://educationnext.org/journal/summer02>.
- 25 Metz, M. (1991). Real School: A Universal Drama Amid Disparate Experience. In D. Mitchell & M. Goetz (Eds.), *Education Politics for the New Century*, 75-91. New York: Falmer.
- 26 Lee, O. & Luykx, A. (2005). Dilemmas in scaling up innovations in elementary science instruction with nonmainstream students. *American Educational Research Journal*, 42(3), 411-438.
- 27 Levin, B. (2012). To know is not enough. *Education Review* 1(1);
- Bryk, A., Gomez, L. & Grunow, A. (2011). *Getting ideas into action: Building networked improvement communities in education*. Stanford, CA: Carnegie Institute for the Improvement of Teaching. Retrieved March 8, 2013, from <http://www.carnegiefoundation.org/elibrary/getting-ideas-action-building-networked-improvement-communities-in-education>.
- 28 Cuban, L. (2008). The perennial reform. *Phi Delta Kappa*, 90(4), 240;
- Slavin, R. E. (1989). PET and the pendulum: Faddism in education and how to stop it. *Phi Delta Kappan*, 70(10), 752-758;
- Levin, B. (2012). Innovation, transformation and improvement in school reform. In C. Chapman, P. Armstrong, A. Harris, D. Muijs and D. Reynolds (Eds.). *School effectiveness and improvement research policy and practice: Challenging the orthodoxy?* (220-229). London: Routledge.
- 29 Dobbie, W. and Fryer, R. (2009). *Are High Quality Schools Enough to Close the achievement Gap? Evidence from a Social Experiment in Harlem* (NBER Working Paper 15473). Cambridge, MA: National Bureau of Economic Research.
- 30 Dobbie, W. and Fryer, R. (2009). *Are High Quality Schools Enough to Close the achievement Gap? Evidence from a Social Experiment in Harlem* (NBER Working Paper 15473). Cambridge, MA: National Bureau of Economic Research.
- 31 Croft, M. & Whitehurst, G. J. (2010, July 20). *The Harlem Children's Zone, Promise Neighborhoods, and the Broader, Bolder Approach to Education*. Washington, DC: Brookings Institution, Brown Center on Education Policy. Retrieved March 10, 2013, from <http://www.brookings.edu/research/reports/2010/07/20-hcz-whitehurst>.
- 32 Dobbie, W. and Fryer, R. (2009). *Are High Quality Schools Enough to Close the achievement Gap? Evidence from a Social Experiment in Harlem* (NBER Working Paper 15473). Cambridge, MA: National Bureau of Economic Research.
- 33 ABC World News (2007, October 17). KIPP officials answer your questions. Retrieved March 10, 2013, from <http://abcnews.go.com/WN/story?id=3741224&page=1>.
- 34 KIPP Foundation (2012). *KIPP new teacher programs*. San Francisco: Author. Retrieved March 11, 2013, from <http://www.kipp.org/files/dmfile/KIPPNewTeacherPrograms2012-13.pdf>.
- 35 Wikipedia (2013). Teach for America. (sections Geographic Reach and Organizational Growth). Retrieved March 11, 2013, from http://en.wikipedia.org/wiki/Teach_For_America.

- 36 Teach for America (2012). *Who we are*. Retrieved March 10, 2013, from <https://www.teachforamerica.org/why-teach-for-america/who-we-look-for/who-we-are>.
- 37 Kopp, W. (2011). *Teach for America, 2011 annual letter*. San Francisco: Teach for America. Retrieved March 11, 2013, from http://www.teachforamerica.org/sites/default/files/Annual.Report.FINAL_.pdf.
- 38 Moore, K. A., et al. (2009, April). *Children in poverty: Trends, consequences, and policy options* (issue brief # 2009-11) Washington, DC: Child Trends. Retrieved March 13, 2013, from http://www.childtrends.org/files/child_trends-2009_04_07_rb_childreninpoverty.pdf.
- 39 Harlem Children's Zone (2009). *History*. Retrieved March 11, 2013, from <http://hcz.org/index.php/about-us/history>.
- 40 Smrekar, C. E., & Mawhinney, H. B. (1999). Integrated services: Challenges in linking schools, families, and communities. In J. Murphy & K. Seashore Louis (Eds.), *Handbook of Research on Educational Administration* (443-461). San Francisco: Jossey-Bass.
- 41 Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, 13(2), 155-190.
- 42 *Washington Post* editorial board (2010, July 20). The time to learn: KIPP schools show what a longer school day offers. *Washington Post*. Retrieved March 11, 2013, from <http://www.washingtonpost.com/wp-dyn/content/article/2010/07/19/AR2010071904357.html>.
- 43 Education Sector (2009, November). *Growing pains: Scaling up the nation's best charter schools*, 10. Washington, DC: Author. Retrieved March 11, 2013, from <http://www.educationsector.org/publications/growing-pains-scaling-nations-best-charter-schools>.
- 44 KIPP Foundation (2011). Independent Auditors' Report and Consolidated Financial Statements (June 2010 and 2011), 3. San Francisco: KIPP Foundation.
- 45 National Center for Education Statistics (2010). *Fast Facts: Teacher trends*; data Derived from U.S. Department of Education, National Center for Education Statistics (2010). *Teacher Attrition and Mobility: Results from the 2008-09 Teacher Follow-up Survey* (NCES 2010-353). Washington, DC: Author. Retrieved March 11, 2013, from <http://nces.ed.gov/fasTfActs/display.asp?id=28>.
- 46 Yasin, Said (1999). The supply and demand of elementary and secondary school teachers in the United States. *Eric Digest*, ED436529. Washington, DC: ERIC Clearinghouse on Teaching and Teaching Education. Retrieved March 11, 2013, from <http://www.ericdigests.org/2000-3/demand.htm>.
- 47 Goetz, M. & Aportela, A. (2008). *Strategic management of human capital: Teach for America*. Madison, WI: Consortium for Policy Research in Education.
- 48 Feistritz, C.E. (2011). *Profile of teachers in the U.S. 2011*, 11, Table 1; 19, Chart 10. Washington, DC: National Center for Education Information. Retrieved March 12, 2013, from <http://www.edweek.org/media/pot2011final-blog.pdf>.
- 49 Carroll, T. (2007). *Policy brief: The high cost of teacher turnover*. Washington, DC: National Commission on Teaching and America's Future;

Milanowski, A., & Odden, A. (2007). *A new approach to the cost of teacher turnover*. Working paper 13. Seattle, WA: Center on Reinventing Public Education.

50 National Center for Education Statistics (n.d.) Table 196.: Enrollment, staff, and degrees/certificates conferred in postsecondary institutions participating in Title IV programs, by level and control of institution, sex of student, type of staff, and type of degree: Fall 2009 and 2009-10. Washingtonm DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved March 13, 2013, from [NCES nces.ed.gov/programs/digest/d11/tables/dt11_196.asp](http://nces.ed.gov/programs/digest/d11/tables/dt11_196.asp).