the evaluation exchange

A PERIODICAL ON EMERGING STRATEGIES IN EVALUATION

HARVARD FAMILY RESEARCH PROJECT ■ HARVARD GRADUATE SCHOOL OF EDUCATION ■ VOLUME XV NUMBER 1 ■ SPRING 2010

from the director's desk

Heather Weiss

Several years ago I visited India and learned about Gyan Shala, a nonprofit start-up that developed an entrepreneurial low-cost model for providing high-quality education to poor children. It was fascinating to hear the story of how this model scaled up over time, starting first with local implementation at the elementary grades, moving next to middle school grades, and finally spreading nationally, as Gyan Shala worked with the government to integrate the model into the public education system. During my career, I have seen numerous examples of programs that were successful as research and demonstration efforts, but then died when they attempted to scale up. The experience of Gyan Shala made me want to take another look at what we know about successfully taking an effort to scale.

Scaling up involves sharing something that is effective (such as programs, practices, or ideas) so that more people can experience its benefits. When something works well, the natural inclination is to share it. The challenge becomes how to do so effectively—and that is where evaluation comes in.

This issue of *The Evaluation Exchange* explores the promising practices and challenges associated with taking an enterprise to scale, along with the role that evaluation can and should play in that process. It is the second in our "hard-to-measure" series, which we inaugurated with our Spring 2007 issue on evaluating advocacy.

Surprisingly few examples exist of nonprofit efforts that have scaled up and achieved lasting success. A program or approach may be strong and effective in one location, but that does not mean it will work the same way in another. Scaling is a complex process that plays out without a script. But we do know that when we take something to scale, we need to start with a clear sense of what is being scaled, why it is being scaled, how the process will work, and what it should look like in the end. This issue of *The Evaluation Exchange* helps readers think through some of those questions and options.

Several articles in the issue make it clear that evaluation is integral to the scaling process. At the beginning of a scaling effort, for example, evaluation can help determine whether something is ready to go to scale and which of its components should be scaled. This stage includes identifying and assessing the "ingredients" that must be in place to successfully scale.

In addition, this issue discusses how to evaluate an effort during the scaling process. Scaling takes place in developmental stages, and information needs differ over time. Evaluation questions and methods must recognize and assess those developmental stages accordingly.

Other articles consider the lessons that experienced programs and evaluators have learned about the scaling process. These lessons draw on successes and failures alike and are offered to help others navigate the process.

Finally, the issue addresses how to evaluate the different approaches that nonprofits use to take something to scale. Several options exist, each with different implications for evaluation.

We hope that this issue of *The Evaluation Exchange* contributes to the conversation about scale in the nonprofit sector and, in particular, reminds us of the important role that evaluation can play in its success.

Heather

Heather B. Weiss, Ed.D.
Founder & Director
Harvard Family Research Project

in this issue

Scaling Impact

Theory & Practice
Broadening the perspective 2
Six steps to scale impact 4

Beyond Basic Training
Five meanings of scale 7
SCALERS model 8

Evaluations to Watch
Save the Children 9

Ask the Expert
Lessons about scale 10

Questions & Answers
A conversation with Mike Smith 12

Promising Practices
Evaluation to support scale 14
Integrated school evaluation 15
WINGS performance data 16
A broader concept of scale 17
RALLY scaling approach 18

Spotlight
Office of Social Innovation 19
eNonprofits study 20
Why Facebook matters 21

New & Noteworthy 23

End Notes
Scaling impact 24

the
evaluation exchange

Founder & Director
Heather B. Weiss, Ed.D.

Managing Editors
Julia Coffman
Erin Harris

HFRP Contributors
Heather B. Weiss, Ed.D.
Julia Coffman
Erin Harris
Priscilla M. D. Little
Heidi Rosenberg, Ph.D.
Helen Janc Malone
Katie Chun

Communications Manager
Marcella Franck

Publications Manager
Carly Bourne

Publications Editors
Judith Maas
Naomi Stephen

Published by
Harvard Family Research Project
Harvard Graduate School of Education

Harvard Family Research Project gratefully acknowledges the support of the Charles Stewart Mott Foundation and the Ewing Marion Kauffman Foundation. The contents of this publication are solely the responsibility of Harvard Family Research Project and do not necessarily reflect the views of our funders.

The Evaluation Exchange accepts query letters for proposed articles. See our query guidelines on our website at www.hfrp.org/evaluation/the-evaluation-exchange/editorial-guidelines

To request a free subscription to *The Evaluation Exchange*, subscribe online at www.hfrp.org/subscribe.

> theory & practice

Broadening the Perspective on Scale

Julia Coffman of Harvard Family Research Project and the Center for Evaluation Innovation describes four approaches to scale that differ on both what is scaled and how it is scaled.

A good deal of the literature on scaling focuses on a definition that refers to the replication (often with room for adaptation) of programs in new sites or locations. But the "what" of scaling can go well beyond programs, and the "how" of scaling can include much more than replication.

In the Winter 2009 issue of the journal *New Directions for Evaluation,* editors Judith Ottoson and Penelope Hawe[1] offer an excellent overview of five theories on how knowledge moves among people and institutions for the purpose of effecting change. These theories have direct relevance to the concept of scale because they describe how things (in this case, knowledge) move or spread—a concept at the heart of scale. The theories help clarify different conceptualizations of what can be scaled, how it can be scaled, and what needs to be considered when evaluating the scaling process.

Adapting content from the *New Directions for Evaluation* journal issue, the matrix on page 3 outlines four definitions of what can be scaled. Programs are one option; others include ideas or innovations, technologies or skills, and policies.

Programs are packaged systems of services that work together to produce impacts for individuals or communities. Before they are scaled, programs usually go through a rigorous evaluation to prove that they are effective. Scaling then involves launching a program in other sites. That process may allow programs to be adapted for different contexts or populations, but adaptations generally do not venture too far from the original model.

Ideas or innovations are new ways of thinking about or doing something. Unlike programs, they are more conceptual and can be approached in different ways. And whereas programs typically come with ownership or branding, ideas generally are not owned by a single individual or organization. An example of an idea is microcredit, the idea of giving very small loans to people in poverty to spur entrepreneurship and self-employment. Microcredit can be approached in different ways, but the core concept is consistent—a small amount of money can make a big difference in people's lives. Scaling ideas involves communicating and spreading them throughout a population or region.

Technologies or skills are products, tools, techniques, or practices. Scaling means increasing the number of people who use them. The scaling process takes place through training and skill development or through marketing and distribution. Examples include training health workers in a new treatment practice, or spreading the use of mosquito nets as a way to fight malaria.

Policies are codified statements about a course of action. Scaling refers to the implementation of those policies throughout an area or a jurisdiction. An example is the statewide adoption and implementation of new quality standards for early childhood programs.

As Ottoson and Hawe demonstrate, conceptualizations of the scaling process go well beyond the replication of programs. When those interested in the scaling process encounter effective practice, they should carefully examine what should be scaled and how. Is it the entire program? Is it the general idea, which can be adapted? Is it certain technologies or skills? Is it policy that codifies and mandates the action that can lead to positive change? Or is it some combination of all four?

Julia Coffman
Senior Consultant, Harvard Family Research Project
Director, Center for Evaluation Innovation
Email: jcoffman@evaluationexchange.org

1. Ottoson, J. M., & Hawe, P. (2009). Knowledge utilization, diffusion, implementation, transfer, and translation: Implications for evaluation. *New Directions for Evaluation, 124,* 1–108. The journal issue covers five "knowledge-for-action" theories: (1) knowledge utilization, (2) diffusion, (3) implementation, (4) transfer, and (5) translation.

FOUR APPROACHES TO SCALE

What Is Scaled	Definition of Scale	Scaling Mechanisms	Relevant Theory/ Literature	Influencing Variables
Program				
A system of projects or services that meets a need for individuals or communities.	Copying a program that research has shown to be effective, with the expectation that it can or will produce the same results in different places. Scaled programs often allow for flexibility in implementation to best adapt to the local context.	• Replication • Adaptation	• Program Theory • Theory of Change	• Fidelity of Implementation • Generalizability • Adaptability
Idea or Innovation				
A new way of thinking about or doing something; new solutions to problems.	Spreading an idea among individuals or organizations within a certain area or system (geographic, organizational, professional); ideas can be adapted to fit different purposes or contexts.	• Communication • Marketing • Dissemination	• Diffusion of Innovation • Tipping Point • Communications Theory	• Relative Advantage • Compatibility • Complexity • Trialability • Observability
Technology or Skill				
Products, tools, techniques, or practices.	Increasing the number of people or places that use or apply a technology, practice, or approach.	• Marketing • Distribution • Training • Granting	• Knowledge Transfer • Technology Transfer • Learning Theory	• Resources • Ability • Market Demand • Existence of Supporting Policies • Availability of Technical Support
Policy				
Codified statements that define plans or a course of action.	Ensuring that ideas expressed as policy are transformed into behavior throughout a place or jurisdiction (e.g., city, county, state, region, country).	• Implementation	• Agenda Setting • Policy Implementation Theory	• Sociopolitical Factors • Governance • Enforcement • Accountability • Resources

Adapted from Ottoson, J. M., & Hawe, P. (2009). Knowledge utilization, diffusion, implementation, transfer, and translation: Implications for evaluation. *New Directions for Evaluation, 124*, 1–108.

Selected Resources on Theory

Program

Fulbright-Anderson, K., Kubisch, A., & Connell, J. (Eds.). (1998). *New approaches to evaluating community initiatives: Theory, measurement, and analysis*. Washington, DC: The Aspen Institute.

Rogers, P. J., Hacsi, T. A., Petrosino, A., & Huebner, T. A. (Eds.). (2000). Program theory in evaluation: Challenges and opportunities. *New Directions in Evaluation, 87,* 1–112.

Idea or Innovation

Gladwell, M. (2002). *The tipping point: How little things can make a big difference*. New York: Back Bay Books.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Technology or Skill

Bozeman, B. (1988). Evaluating technology transfer and diffusion. *Evaluation and Program Planning, 11*(1), 63–104.

Goh, S. C. (2002). Managing effective knowledge transfer: An integrative framework and some practice implications. *Journal of Knowledge Management, 6*(1), 23–30.

Policy

Goggin, M. L., Bowman, A. O., Lester, J. P., & O'Toole, L. J. (1990). *Implementation theory and practice: Toward a third generation*. Glenview, IL: Scott Foresman.

Sabatier, P. A. (2007). *Theories of the policy process* (2nd ed.). Boulder: Westview Press.

Six Steps to Successfully Scale Impact in the Nonprofit Sector

Erin Harris of Harvard Family Research Project discusses how nonprofits can successfully scale up an intervention, based on a review of the literature on this topic.

To increase their impact, many nonprofits seek to "scale up" or "go to scale" by expanding their interventions to reach larger populations. This scaling process most commonly involves implementing the intervention at new sites or expanding the capacity of existing sites to serve a larger number of participants.

So why go to scale? For nonprofits, the primary goal of scaling is to spread the impact of the intervention. Scaling can also be a practical approach: Rather than starting from scratch, it often makes more sense to take a model that works well in one community and try it out in another. Interventions that go to scale can also benefit from being part of a larger network where they can share resources and operating procedures. This larger network can often allow an intervention to produce bigger outcomes and at a faster pace than each individual site would be able to do on its own. Further, demonstrating impact on a larger scale can help create greater visibility and leverage in attracting additional support for the intervention.

Rather than starting from scratch, it often makes more sense to take a model that works well in one community and try it out in another.

Despite these benefits, many interventions fail to successfully go to scale. The process of scaling is not easy; it requires stakeholder support and buy-in, careful planning and assessment, and sufficient resources to maintain quality. This article describes the steps that nonprofits should follow in going to scale to help ensure success. These recommended steps are based on a review of the literature on scale (including the articles in this *Evaluation Exchange* issue).

1. *Determine whether the intervention is ready to go to scale.*

There are several issues to consider in determining whether an intervention is ready for the scaling process. First and foremost, the intervention must have a strong theory of change or logic model that links the program inputs to the outcomes. This theory of change or logic model should convey which components are essential to the intervention's success and which can allow for flexibility so that they can be tailored to local needs and available resources.

Second, the nonprofit needs to test the theory of change or logic model through evaluation of program implementation and outcomes to determine: (a) whether the intervention was implemented as intended, (b) whether the intervention was effective, (c) why the intervention was effective (i.e., what elements of the intervention were linked to its success), (d) how the elements that were linked to success can be transferred to new settings, and (e) whether the intervention's effects were large enough to persist in the face of contextual variability.

Third, the intervention must have buy-in from stakeholders (including funders) for the scaling process. It is crucial that the intervention have support, especially in the form of additional funding and other necessary resources to ensure that the quality of the services and staff can be maintained at additional sites. Nonprofits can use the results of the evaluation to help convince funders and other key stakeholders of the value of the program and of how these benefits can be spread to new sites.

2. *Select the best approach to bring the intervention to scale.*

The term "scale" is used in the nonprofit sector to describe different approaches for expanding an intervention's impact. Though there is no one right way to scale, it is important for nonprofits to consider all of the possible approaches for going to scale and then choose the one that best meets their needs.

In the business world, bringing something to scale usually refers to replicating something (e.g., a business model or service) in a new site with as much fidelity to the original as possible. As applied to the nonprofit sector, however, replication tends to have a more nuanced meaning, where the goal is to replicate the program's impact rather than its components exactly as they were initially implemented. Replication is the most straightforward way of conceptualizing scale and allows nonprofits to benefit from the lessons from the business sector. However, replication as it applies to the nonprofit sector has been criticized for its narrow scope, in not viewing scale beyond the goal of growth in numbers served.

Another approach that has grown in popularity in the nonprofit sector is to scale ideas, regardless of whether those ideas include the "branding" of the original intervention. This broader view of scale takes into account the fact that ownership of the initiative is less of a motivating factor in the nonprofit sector than in the business world. It is often discussed using the term, "the traveling of ideas," meaning that promising practices are communicated to others in a given field to help "diffuse new concepts, processes, and techniques."[1] This approach allows for greater adaptability to different settings and contexts.

A third model of thinking about scale in the nonprofit sector is to view it as having multiple facets, with various ways of scaling an intervention's impact. For example, Cynthia Coburn describes scale as including four components: spread, depth, sustainability, and shift in reform ownership (see the article, *Applying a Broader Concept of Scale to Evaluate a Funding Strategy,* on page 17).

1. Scott, W. R., Deschenes, S., Hopkins, K., Newman, A., & McLaughlin, M. (2006). Advocacy organizations and the field of youth services: Ongoing efforts to restructure a field. *Nonprofit and Voluntary Sector Quarterly, 35* (1), 691–714.

Similarly, Peter Frumkin describes scale as having five different meanings: financial strength, program expansion, comprehensiveness, multisite replication, and accepted doctrine (see the article, *The Five Meanings of Scale in Philanthropy,* on page 7). Both of these models help get at the complexities of scale in the nonprofit sector and create a conception of scale that is more meaningful and sustainable than simple replication.

3. *Select sites that are best suited to the intervention.*

As indicated above, the intervention's theory of change should provide information about what elements of the intervention are necessary for its success and where there is room for negotiation and flexibility. For example, in selecting new sites to implement the Nurse-Family Partnership program, the program developer, David Olds, determined that the new sites had to have "certain capacities to operate and sustain the program with high quality." In particular, the sites had to have the capacity for "an organization and community that are fully knowledgeable and supportive of the program; a staff that is well trained and supported in the conduct of the program model; and real-time information on implementation of the program and its achievement of benchmarks to guide efforts in continuous quality improvement."[2] As part of the selection process, nonprofits should ensure that the communities being considered for the new sites have the local capacity to support the elements of the intervention that are necessary for its success.

It is crucial that the intervention have support, especially in the form of additional funding, to ensure that the quality of the services and staff can be maintained at additional sites.

In addition, the intervention should be implemented in communities where there is an identified need that is not being sufficiently met by existing community resources. Thus, the site-selection process should include ensuring that the program is not moving into an area that is already saturated with similar services. Otherwise, the intervention is likely to find itself competing for participants and local resources (e.g., quality staff) while adding little value to the local community.

4. *Develop the capacity and infrastructure to manage multiple sites.*

Managing multiple sites requires additional structures and procedures to ensure that these programs are effectively and efficiently managed across the larger system. Several pieces must be in place to ensure that the organizational infrastructure can support the scaling process.

First, nonprofits need to develop management structures to coordinate services both across and within sites. In particular, the intervention must develop and communicate roles for local sites and central offices so that it is clear who is responsible for which aspects of running the program. Systems must be in place for communication across the various levels so that everyone is working toward a common set of goals as part of a coordinated team effort.

Second, resources and structures must be in place at the local level that can help ensure quality, such as strong local leadership and qualified service delivery staff. This piece often proves to be one of the most challenging issues for nonprofits in scaling: how to maintain quality as the program grows. However, if the sites have been carefully selected as those best suited to the intervention's theory of change, quality control will be less of a problem. Still, program developers need to be vigilant in ensuring that quality standards are met and upheld. To ensure quality, nonprofits need to allocate resources to evaluate implementation and outcomes across sites (see step 5 below for more detail).

Third, program leaders should be aware and take advantage of resources and strengths in the local community that can benefit the intervention. At the same time, program developers should be aware of deficits in the local setting that they will need to overcome (or accommodate). To deal with the local context, flexibility in implementation can help strengthen the intervention. For example, an after school program being implemented in a community with a strong school system might consider finding ways to partner with local schools to hire school-day teachers as program staff. That same program might also want to work with other local after school programs to coordinate their services to complement (rather than duplicate) one another. Strategies will differ from site to site, depending on the strengths and weaknesses of the local context.

5. *Evaluate the scaling process.*

As the intervention is implemented at new sites, the scaling process itself should be evaluated. Implementation evaluation conducted during the scaling process should focus on continuous learning and improvement; as part of this process, the evaluation findings can help inform any necessary adjustments that need to be made to the program to account for the local context. These adjustments are often inevitable; even with a strong theory of change, it can be difficult, if not impossible, to anticipate all the possible ways in which the local context may affect program implementation and outcomes.

Evaluation can also take place at the end of the scaling process to determine the successes (and failures) of the process and what led to those successes and failures. Nonprofits can then use these lessons to help inform the intervention's future scaling strategies and to provide guidance to other interventions in their scaling processes. They can also use positive outcomes to promote the intervention to potential new stakeholders (including funders)

2. Olds, D. L. (2007). *The Nurse-Family Partnership: From trials to practice*, p. 27. Federal Research Paper. Online at: www.earlychildhoodrc.org/events/presentations/olds.pdf. For more information on the scaling process of the Nurse-Family Partnership, see: Goodman, A. (2006). *The story of David Olds and the Nurse Home Visiting Program*. Princeton: The Robert Wood Johnson Foundation. Online at: www.rwjf.org/files/publications/other/DavidOldsSpecialReport0606.pdf.

and to demonstrate success to existing stakeholders. Thus, these positive outcomes can be helpful in generating additional funding and other resources to scale the intervention even further.

6. *Share promising practices and lessons about scale with other nonprofits.*

Scale-up is a relatively new concept in the nonprofit sector. Theories and practices related to scale-up in the nonprofit world are still emerging as those involved work to identify best practices. Although some of the frameworks and lessons from scale can be adapted from the business world, the nonprofit world has its own set of opportunities and challenges that affect the scaling process, not the least of which is the fact that the motivations for scaling are often less straightforward than in the business world, where profit is usually the primary motive. The more nonprofits are willing to take the risk to go to scale, and work to overcome the challenges in doing so, the more opportunities will be available to identify and adapt promising practices in the scaling process specific to the needs of nonprofits.

Erin Harris
Project Manager, Harvard Family Research Project
Email: erin_harris@harvard.edu

Key Readings on Scale

Bradach, J. (2003, Spring). Going to scale. *The Stanford Social Innovation Review*, 18–25. Online at: www.ssireview.org/articles/entry/going_to_scale/.

Campbell, K., Taft-Pearman, M., & Lee, M. (2008). *Getting replication right: The decisions that matter most for nonprofit organizations looking to expand*. New York: Bridgespan Group. Online at: www.bridgespan.org/LearningCenter/ResourceDetail.aspx?id=398.

Carlson, N. F. (2008). *Replicating success: A funder's perspective on the "why" and "how" of supporting the local office of an expanding organization. A case study of Blue Ridge Foundation New York's support of the Taproot Foundation NYC*. Brooklyn: Blue Ridge Foundation New York. Online at: www.brfny.org/downloads/BlueRidge-TaprootFoundationCaseStudy.doc.

The Center for the Advancement of Social Entrepreneurship. (2006). *Scaling social impact research project thought leader meeting annotated bibliography, v. 1.0*. Durham, NC: Duke University Fuqua School of Business. Online at: www.caseatduke.org/knowledge/scalingsocialimpact/bibliography.html.

Clarke, J., Dede, C., Ketelhut, D. J., & Nelson, B. (2006). A design-based research strategy to promote scalability for educational innovations. *Educational Technology, 46*(3), 27–36. Online at: http://muve.gse.harvard.edu/rivercityproject/documents/final_scale_article_Clarke.pdf.

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12. Online at: http://gse.berkeley.edu/faculty/CECoburn/coburnscale.pdf.

Creech, H. (2008). *SEED Research and Learning Programme 2007–8: Scale up and replication for social and environmental enterprises*. Gland, Switzerland: The SEED Initiative. Online at: www.seedinit.org/download-document/13-seed-iisd-report-on-scale-up-and-replication-280708sml.html.

Edna McConnell Clark Foundation. (2008). *An experiment in coordinated investment: A progress report on the Edna McConnell Clark Foundation's growth capital aggregation pilot*. New York: Author. Online at: www.emcf.org/pdf/gcap_progressreportOct08.pdf.

Elmore, R. E. (1996, Spring). Going to scale with good educational practice. *Harvard Educational Review, 66*(1), 1–27.

Farole Jr., D. J. (2006.) *The challenges of going to scale: Lessons from other disciplines for problem-solving courts*. New York: Center for Court Innovation. Online at: www.courtinnovation.org/_uploads/documents/Lessons.pdf.

McDonald, S. (2009). Scale-up as a framework for intervention, program, and policy evaluation research. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 191–208). New York: Routledge Publishers (for the American Educational Research Association).

McLendon, L. L., & Polis, K. (2007). *The state adult education staff guide for going to scale: A guide for planning, implementing, and evaluating program improvement initiatives*. Washington, DC: The National Adult Education Professional Development Consortium. Online at: www.naepdc.org/GTScorrected3.11.doc.

Summerville, G., & Raley, B. (2009). *Laying a solid foundation: Strategies for effective program replication*. Philadelphia: Public/Private Ventures. Online at: www.ppv.org/ppv/publications/assets/298_publication.pdf.

Uvin, P., & Miller, D. (1996). Paths to scaling-up: Alternative strategies for local nongovernmental organizations. *Human Organization, 55*, 344–354.

The Five Meanings of Scale in Philanthropy

Peter Frumkin of the University of Texas at Austin describes the five primary ways in which funders define scale as it relates to nonprofits' efforts to create a lasting and significant impact.

For nonprofit organizations, being effective means more than simply carrying out an initiative successfully and meeting the needs of a small group of people. Effectiveness also involves reaching many people—taking the social leverage that an intervention creates and amplifying it even more broadly. From a philanthropic perspective, scale means investing in programs and initiatives run by nonprofit organizations that are able to create a lasting and significant impact. Donors speak variously of "taking a program to scale," "going to scale," and "scaling up." No matter how they refer to it, foundations define a nonprofit organization's ability to go to scale using five dimensions: financial strength, program expansion, comprehensiveness, multisite replication, and accepted doctrine.

Financial strength refers to a nonprofit's organizational strength and sustainability, often secured by endowment or operating budgets with dependable revenue streams. In a sector in which financial crises are commonplace, scale means being able to ride out the storms. Nonprofits that have gone to scale under this definition remain few; those that have are concentrated in categories of nonprofits such as museums and universities.

Program expansion describes the scope of service, usually measured by the number of clients served. When a program demonstrates positive outcomes for its participants, the goal is often to provide more funding and bring it to more people. To a funder, enabling an organization with a proven track record to expand its operations represents a high-return, low-risk activity.

As soon as an initiative seems to achieve significant results, one of the first impulses of nonprofit managers and funders alike is to ramp up the effort and find a way to identify and serve more clients. Their ultimate ambition is often to secure local, state, or federal funding after the initiative launches with private money.

Comprehensiveness refers to programs that constitute a coherent set of resources for clients or communities, brought together under one roof to provide an integrated group of activities. This configuration mitigates the coordination problems inherent in the nonprofit sector's division of labor and ability to proliferate. Seeking to create synergies by funding a collection of services, many donors perceive scale as being closely linked to building a sizable local presence and thus having a more fundamental and lasting impact to achieve systemwide reform.

Multisite replication refers to attempts to dissect the essential elements of a successful initiative or service model in order to reconstruct the effort elsewhere with different personnel and under different circumstances. Replication can proceed in two ways: (1) within the organization, through a set of chapters or a franchise system linking independent organizations; or (2) outside the organization, through independent efforts to create similar programs. Although funders may be able to foster some replication through the use of grants and incentives, most externally directed replication efforts will struggle with the vast, unruly, and idiosyncratic tide of nonprofit organizations that resist imitation and convergence. Some innovations and ideas have been replicated, but large numbers of projects are unable to find any takers, even when they have shown great promise. Replication may rest on the shaky assumption that nonprofits are amenable to cookie-cutter duplication.

Accepted doctrine focuses on the power of formulating and diffusing a new and accepted doctrine within a given field. Creating a new doctrine is different from other forms of scale because it seeks to change the conceptual and intellectual frame surrounding a particular field. A successful effort at doctrine building brings about a major shift in a field by changing the ways in which people think about their work and carry out their programs; it creates a wide-ranging and lasting impact by leading to a wholesale reevaluation of a field's standard operating procedures and assumptions.

Achieving scale raises a number of questions: When and why should these scale strategies be applied? Why does scale sometimes fail? Does the ideal of scale fit the business sector better than the nonprofit sector? Can commitments to scale and equity be embraced simultaneously? Are less successful attempts at scale simply examples of domains in which public policy should be allowed to operate?

Most of these questions remain unanswered even though scale has become a major target of philanthropy. Rather than engage these difficult questions, it is far easier to simply fall back on arguments about spreading an initiative's benefits more broadly and achieving efficiencies.

The fundamental problem with the concept of scale, as understood in philanthropy, is the assumption that the scope of public impact correlates with the public value created. Scale seems an obviously desirable objective by virtue of simple math: Assisting 10,000 clients seems better than assisting 1,000. This assumption is particularly hard to defeat because it is notoriously difficult to measure the quality of nonprofit programs, whereas it is easy to use number served as a proxy for impact. Strategic giving requires a nuanced and guarded stance with regard to scale, one that is grounded in a clear understanding of the many meanings—and limits—of scale.

Peter Frumkin
Professor of Public Affairs and Director
RGK Center for Philanthropy and Community Service
Lyndon B. Johnson School of Public Affairs
The University of Texas at Austin
Email: frumkin@mail.utexas.edu

Related Resources

The definitions of scale discussed in this article are presented in greater detail on the Social Edge blog at: www.socialedge.org/blogs/frumkin-on-philanthropy and in Frumkin, P. (2006). *The art and science of philanthropy*. Chicago: University of Chicago Press.

Scaling Social Entrepreneurial Impact: The SCALERS Model

Paul N. Bloom and Aaron K. Chatterji of Duke University discuss a model that they have developed to conceptualize scaling impact for social entrepreneurs.

Scaling social impact has become a major challenge for social entrepreneurs—individuals who start up and lead new organizations or programs to address social problems using change strategies that differ from those used in the past. Social entrepreneurs who achieve initial success with their ideas often have difficulty replicating these ideas on a larger scale.

To determine what factors contribute to success when social entrepreneurs scale up their efforts, we developed a model that identifies seven organizational capabilities, or "drivers." We also indicate situational contingencies that might lead some drivers to be more effective in certain situations than in others. The drivers are identified by the acronym "SCALERS": staffing, communicating, alliance building, lobbying, earnings generation, replicating, and stimulating market forces.

Staffing refers to the organization's effectiveness at filling its labor needs—including managers, staff, and volunteers—with people who have the right skills for their positions. Organizations that prioritize this driver pay close attention to their personnel and human resource functions, so that recruiting, training, appraising, and compensating staff are done competently. Many organizations must pay equal or greater attention to recruiting, training, and managing unpaid volunteers, who are often the lifeblood of cash-starved social organizations. Boards of directors need to be adept at identifying, recruiting, guiding, and retaining top management to lead the organization.

Communicating refers to the organization's ability to convince stakeholders that its strategy is worth adopting or supporting. Placing a priority on this driver means the organization is successful at persuading potential beneficiaries to take advantage of its services or to change their behaviors in socially beneficial ways, convincing volunteers and employees to work for the organization, encouraging consumers to patronize the organization's income-generating activities, persuading donors to provide funds to the organization, or creating favorable public attitudes toward the organization's programs.

Alliance building is the effectiveness with which the organization has forged partnerships and other linkages to bring about desired social changes. Recent research has identified alliance building as an essential ingredient in successful scaling. Organizations that employ this capability effectively do not try to do things on their own, but instead forge unified efforts.

Lobbying is the organization's ability to advocate for government actions that may work in its favor.[1] Organizations that lobby skillfully succeed in getting courts, administrative agencies, legislators, and government leaders to help their cause.

Earnings generation refers to the effectiveness with which the organization generates a stream of revenue that exceeds its expenses. Organizations that are successful at generating earnings do not have trouble paying their bills or funding their activities.

Replicating refers to an organization's effectiveness in reproducing its programs and initiatives. An organization that is adept at replication ensures that its services, programs, and other efforts can be copied or extended without a decline in quality; training, franchising, contracting, and other tools are used to ensure quality control.

Stimulating market forces covers an organization's ability to create incentives that encourage people or institutions to pursue private interests while also serving the public good. An organization with this capability is successful at creating markets for offerings (i.e., products and services) such as microloans or carbon credits. Stimulating market forces can lead to significant social change.

The SCALERS model proposes that the extent to which an individual driver influences scaling success depends on various factors in an organization's internal and external environment that can enhance or suppress a driver's influence. For example, the degree to which staffing drives scaling success depends on the organization's labor needs. When labor needs are high (e.g., the organization provides highly skilled services to large numbers of clients), staffing is crucial for successful scaling. However, when labor needs are low (e.g., the organization's strategy is not based on service delivery), other drivers are more critical to success. In some situations, effective deployment of all seven drivers is needed for successful scaling. In others, success is dependent on a few critical drivers.

Taken as a whole, the SCALERS model offers a roadmap for social entrepreneurial organizations interested in scaling their impact. In addition, it can act as an evaluation framework that helps social entrepreneurs to track and assess scaling progress and to identify ways to improve. For example, an organization might use the model to assess its performance and determine how past actions have helped or hurt its ability to scale. The model can also appraise the potential impact of future plans. And, finally, an organization might use the model to examine its ecosystem and determine how situational contingencies affect the seven drivers, both negatively and positively.

Paul N. Bloom
Adjunct Professor of Social Entrepreneurship and Marketing
Email: paul.bloom@duke.edu

Aaron K. Chatterji
Assistant Professor
Email: ronnie@duke.edu

Fuqua School of Business
Duke University

1. The term "lobbying" is used loosely here and does not refer only to registered lobbyists, which could jeopardize an organization's tax-exempt status.

Related Resources

Bloom, P. N., & Chatterji, A. K. (2009). Scaling social entrepreneurial impact. *California Management Review, 51*(3), 114–133.

Bloom, P. N., & Chatterji, A. (2009). *Scaling social entrepreneurial impact*. Boston: Harvard Business Publishing.

Save the Children's Literacy Programs in Rural America: Evaluation That Informs Scale-Up

Elizabeth Reisner of Policy Studies Associates discusses how the learning gains of a children's literacy program relate to the program's scaling process.

A major goal of Save the Children (STC) in the United States is to improve the literacy skills of children who live in the nation's poorest rural communities, including locations in Appalachia, the Southeast, the Mississippi River Delta, the Gulf Coast, the Southwest, and California's Central Valley.

The organization's literacy program began in 2003 with 1,800 children in 15 pilot sites. These sites implemented the program's after school literacy model, which at that time emphasized guided independent reading practice and employed leveled books (i.e., books organized by their degree of difficulty), computer-assisted quizzes (i.e., quizzes administered on a computer), and periodic assessment using a standardized, norm-referenced reading test.

Also in 2003, STC hired Policy Studies Associates, a Washington, DC research firm, to launch a program evaluation. STC asked us to measure program participants' change in literacy skills and to analyze these results in light of participants' differences in program attendance, baseline literacy skills, and grade in school.

Because we initiated the evaluation just as STC began program implementation, we had a chance to recommend program features that would facilitate evaluation. In particular, we encouraged STC to design its attendance database to record the number of books that each child read as part of the literacy program and each child's scores on quizzes taken after reading each book. We also helped design the attendance database so that it could be merged with records of each child's scores on the program's reading test.

Using these data, we have produced a series of six annual evaluation reports that describe the baseline educational characteristics of participating children, their program attendance, their level of engagement in the program (measured by books read, quizzes taken, and quiz scores), and their change in reading proficiency (measured in comparison with the reading test publisher's national norming group). We have also produced tailored annual reports for each program site and for clusters of sites with shared characteristics.

By the end of the 2008–09 program year, the initiative had grown to 146 sites serving almost 15,000 children. In addition to guided independent reading practice, the program now includes skill-building tutorials as well as activities to improve fluency and vocabulary, any of which can be offered in school or after school.

Evaluation findings show that 60% of 2008–09 participants increased their literacy performance by at least 2 NCEs (normal curve equivalents) over and above the learning gain that would typically result from another year in school. Participants' average literacy gain was 5.8 NCEs beyond the typical expected annual gain. Among children who participated in the program 55 days or more in 2008–09, the proportion improving by at least 2 NCEs was 63%. Among children who attended 55 days or more and who scored below grade level at the fall 2008 baseline, 66% gained 2 or more NCEs. Higher numbers of books read and quizzes passed were statistically associated with higher NCE gains.

Although these findings are subject to selection bias because children attended the after school sessions voluntarily, they clearly suggest that the literacy program was making a difference for participating children at the same time that it was scaling up. These observations prompt a question: Did evidence of participant learning influence the growth and further development of the program?

In the current tough economic climate, state and local leaders will not consider any new education initiative unless it provides explicit evidence of effectiveness.

For the *Evaluation Exchange,* I asked STC program leaders to describe how program growth and documented learning gain intersected, if at all. Mark Shriver, STC's vice president for U.S. programs, said that the relationship was positive and powerful. When he meets with educators and policy leaders to promote the program, he said, evaluation findings are what he talks about after he says hello. In the current tough economic climate, according to Shriver, state and local leaders will not consider any new education initiative unless it provides explicit evidence of effectiveness. Andrew Hysell, STC's associate vice president for policy and advocacy, said that educators and policy leaders who make program decisions want to see evaluation evidence that is both rigorous and focused on state educational priorities such as literacy.

Ann Mintz, STC education advisor for U.S. programs, has another perspective on the relationship of evaluation and scale-up. For her, evaluation is an essential tool in tracking program effectiveness, determining what works best in the program, and providing informed feedback to improve the program. For example, based on evaluation findings about which children benefit most, STC urges program leaders to recruit and retain children who fit the literacy program's individual success profile (e.g., struggling readers who are capable of regular attendance and active engagement in the program's literacy activities). Because regular program attendance appears to be so important in producing gains, STC encourages programs to educate parents about the value of regular attendance and to create "authentic" ways of encouraging attendance and personal engagement in program opportunities (no pizza parties for this health-conscious organization). According to Mintz, rewards for high attendance and engagement in STC programs include gifts of new books, the chance to read to kindergarten students, and the opportunity to eat lunch with the school principal.

continued on page 22

Lessons from Evaluators' Experiences with Scale

Heidi Rosenberg of Harvard Family Research Project and Helen Westmoreland of the Flamboyan Foundation spoke with several evaluators and program administrators about their experience evaluating the process of going to scale.

Harvard Family Research Project spoke with three evaluators (see box below) to discover how evaluation can inform and assess scaling efforts. They shared lessons from their experiences in evaluating programs as they went to scale. Themes that emerged from these conversations follow.

The ways in which programs think about scale are driven in part by their particular characteristics and activities.

Scaling takes different forms, from replication of an established program model, to a deepening of program efforts to achieve more measurable impact, to scaling components that have demonstrated success. The approach taken depends in large part on the needs and goals of the program being scaled.

Replication. The replication model is traditionally used in the business sector and has been adapted for the nonprofit sector. In the case of Citizen Schools, the organization wished to scale its after school program across the country with the goal of faithfully replicating the original program model. Based on PSA's evaluation of the original program, which demonstrated its positive impact on student achievement, leadership at Citizen Schools felt the program as a whole was ready to expand beyond the home base.

Researchers Interviewed

Elizabeth Reisner, Policy Studies Associates (PSA)

Evaluator of Citizen Schools, which sponsors after school education programs for disadvantaged middle school youth. PSA conducted an impact study evaluation of the original program in Boston and an implementation evaluation of the national program.

William Penuel, Center for Technology, SRI International

Evaluator of the GLOBE Program, an international program in earth science and earth science education. SRI's evaluation focused on the effects of state policies and professional development on the implementation of the science curriculum.

Robin Galloway, Research Institute for Studies in Education, Iowa State University

Evaluator of the Iowa Parent Information and Resource Center (PIRC), a federally funded program that provides regional and statewide services to promote family engagement in education. **Ed Redalen**, Iowa PIRC program director, and **Ron Mirr**, a consultant to the Iowa PIRC, also contributed.

Deepening of program efforts. Another way of framing scaling efforts relates to the depth of a program's impact and reach. The Iowa PIRC team took this approach: Scaling involved reframing the program's efforts from simply engaging in a wide variety of parent involvement activities to focusing on the degree to which the PIRC showed evidence of impact along a set of meaningful outcomes. In this case, scaling related to the *depth* of program impact. The PIRC team reported that its early efforts sought to raise parent awareness, but that program administrators became dissatisfied with the lack of measurable impact. As Mirr commented, "Our scaling went from a shallow understanding of parent involvement activities to committing to change in depth over time."

Identification of core components. Not all elements of a program can, or should, be scaled, since program components are often context-specific and do not translate well to different locations. In these instances, the most promising practices are identified and selectively scaled. SRI's evaluation of the GLOBE Program identified successful professional development practices among widespread, highly variable program sites; these practices were then scaled across the program, thus increasing the number of sites that adopted these components. Penuel noted, "The GLOBE Program had wide variability in the success of its local partners. As part of our evaluation, we went in to help them understand, through comparative case study analyses of successful partners, the factors of professional development that were associated with higher levels of implementation."

Different phases of the scaling process require different evaluation approaches.

To evaluate scale effectively, researchers need to choose methods and approaches based on the stage of the scaling process and determine how the data will be most useful to the program at each stage. During preparation, evaluation can help identify what program elements need to be in place in order to scale. Evaluation efforts may shift depending on the scaling approach selected. Once programs have gone to scale, researchers can use evaluation data to assess how well the program has scaled.

Using evaluation data to decide what elements to scale. Base-level evaluation data can help researchers identify the program elements that must be in place for successful program expansion. Reisner used the initial Boston-area impact study to identify the key components of Citizen Schools that affected student success. Program staff analyzed the contexts and resources necessary to create those core components and built that understanding into a consideration of where to implement program services. SRI's Penuel reported that evaluating widely scaled, field-building program efforts requires identifying the common elements underlying program success: "It is not particularly informative to say simply that the context matters and that the context is messy. If we are thinking about field building in this area, we have to assume that there are some regularities across contexts that matter. For example, it matters that teachers perceive that the professional development and the program that you are trying to implement is coherent with the school and district and state goals for learning." Once researchers have identified those regularities, or common elements, Penuel pointed out, the next step is to ask, "How do I measure that? And how do I fit a model that goes across many schools, in order to try to test my conjecture about the importance of that factor?"

Tailoring evaluation approaches to the scaling strategy. In Iowa, the PIRC evaluation started by measuring program inputs (i.e., the services and resources provided). After deciding to shift their emphasis toward deepening their services, PIRC administrators and the evaluator examined their qualitative data to identify "champion" teams within their Sustaining Parent Involvement Network (SPIN) and targeted those sites for the initial scaling. This shift from counting program inputs to assessing depth of impact required program administrators to work closely with their evaluator; Galloway held frequent meetings with the PIRC team to help its members plan how their proposed scaling efforts and points of impact could be operationalized and evaluated to show effect.

Using evaluation data to assess the scaled program. Once a program has gone to scale, the evaluation activities shift to examining how the scaling process is working. The implementation study of the expanded national Citizen Schools program allowed the organization to assess implementation consistency across sites and to develop a more cohesive sense of program identity. The positive findings from both the impact study of the initial site and the follow-up study of the scaled program encouraged Citizen Schools to consider launching a rigorous experimental design impact study. As Reisner noted, "Our evaluation describes the similarity and diversity across sites, and the news is pretty good in terms of the consistency of implementation nationally. If the program just varied all over the map, then it would be premature to conduct a very rigorous impact evaluation of the program."

If you really want an education program to scale, there is no point at which it becomes less resource-intensive.

Effective data collection systems are crucial to successful scaling and evaluation.

Reisner and the Iowa PIRC team discussed the importance of strong data collection efforts—sites must be willing to provide the necessary data, and evaluators need an appropriate infrastructure to gather and disseminate data. As Reisner observed, "Getting the cooperation of the school system for data collection and research is a nontrivial matter. It is much easier to do it on the front end than to go back and have to fill in after the fact."

Structures and norms for collecting data are important for any evaluation, but are especially critical for scaling efforts. Proper data collection and documentation help lay the foundation to assess a program's progress and reveal which elements of program success should be scaled.

Programs need to ensure that their sites are collecting the same core set of data in a coordinated manner. The Iowa PIRC team's scaling efforts, for example, involved a shift from disparate data collection methods among SPIN sites to procedures that everyone in the network followed. Mirr portrayed it as a "shift from essentially independent or no data to common data that everybody collects the same way."

Scaling requires significant time and resources that may entail compromises or even sacrifices in other areas.

SRI's Penuel noted a tendency to view educational and social service scaling as a process akin to corporate "software start-up" scaling models, where costs are concentrated on development, and dissemination or replication becomes much less costly and resource-intensive over time. "In my view, educational scaling is not at all like software; it is like services. If you really want something to scale, there is no point at which it becomes a whole lot less resource-intensive."

In Penuel's view, this resource-intensiveness creates a tension between the demands of policymakers—whose support can impact programs' funding and scaling opportunities—and the realities of careful, authentic scaling of educational programs: "Policymakers are always looking for ways to scale at low cost and the point at which they are going to be able to completely transfer ownership of the innovation to the field of practice; in only rare cases does that ever happen. And usually it happens in cases where there is an easy fit to the context. And if we are trying to make *big* changes, an easy fit to the context is not going to have a dramatic effect on practice. The things that have the potential to change are game-changers—they push against the main culture of education. They do things really differently and therefore require ongoing efforts to organize."

Noting that scaling requires significant resources, Reisner pointed out that successful evaluations and program scaling efforts can lead to difficult choices. "Citizen Schools is in the process of determining how they are going to move forward in terms of the next phase of impact evaluation." The sequencing of the program's evaluations thus far has, as Reisner put it, "positioned Citizen Schools to be able to make some very difficult choices in their investments. On the one hand, they would probably like to

continued on page 22

Related Resources

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44* (4), 921–958.

Vile, J. D., Arcaira, E., & Reisner, E. R. (2009). Progress toward high school graduation: Citizen Schools' youth outcomes in Boston. Washington, DC: Policy Studies Associates. Available at: www.policystudies.com/studies/youth/Citizen%20Schools%20Phase%20VI%20Report.pdf.

> questions & answers

A conversation with

Mike Smith

Marshall "Mike" Smith is senior counselor to the secretary as well as director of international affairs at the U.S. Department of Education. His previous positions in government have included chief of staff to the secretary for education and assistant commissioner for policy studies in the Office of Education under the Carter administration and undersecretary and acting deputy secretary for education in the Clinton administration. Dr. Smith has been a professor at Harvard University, the University of Wisconsin–Madison, and Stanford University, where he was also dean of the School of Education. Most recently, he was program director for education at the William and Flora Hewlett Foundation. His distinguished career includes positions on many national commissions and panels, and he has authored numerous publications. Dr. Smith earned his master's and doctoral degrees from the Harvard Graduate School of Education.

Q **What does the phrase "going to scale" mean?**

A In the current education environment, going to scale means taking a promising innovation and replicating it in a large number of places. Going to scale at a significant level means spreading an innovation throughout an entire geographic region. In the policy environment, going to scale means taking an idea that seems to work in a particular setting or in multiple settings, codifying it, and then enforcing it through state or federal legislation.

Q **Why did the idea of scale come into the policy conversation in education?**

A Policymakers were frustrated that promising innovations and good ideas were not spreading. I remember President Clinton asking, "Why don't ideas travel?" What he meant was, "Why aren't good ideas replicated in other settings?" Education reforms in this country are like fireflies in a field—the fireflies blink on and off, but they are isolated and uncoordinated, so they do not give off a concentrated or meaningful glow. Also, when things are tried and they fail, we have a tendency to try the same thing over again later. Or we try reforms that work for a while, but we do not support them long term and they go away.

In response to this frustration, there has been a growing sense that we should stop starting from scratch in education reform and build more on what has worked. This is where the idea of scale comes in. If we can learn how to bring good ideas to scale, then we can start to make more progress in reforming education.

Education reforms in this country are like fireflies in a field—the fireflies blink on and off, but they are isolated and uncoordinated, so they do not give off a concentrated or meaningful glow.

Q **How do you know when an intervention is ready to go to scale?**

A For an intervention to go to scale, it must have *external validity*—it must have similar effects in a variety of contexts. There are several considerations in determining external validity.

First, the intervention needs a strong theory of change or logic model that identifies the causal drivers necessary to produce the intervention's outcomes. As part of this model, it is important to identify the contextual variables that can impact, positively or negatively, the success of those causal drivers. For example, an intervention that involves a reduction in school class size will probably not improve student achievement without additional high-quality teachers to support that reduction. A good understanding of the intervention's causal drivers and the environmental conditions required for them to be successful leads to a good understanding of the contexts in which that intervention can and should be applied.

Second, the intervention needs to demonstrate large effects. If the intervention takes place in an "ideal" context, and the effects are relatively small, when the intervention is moved to another location where conditions are not optimal, the effects will be overcome by context. The effects need to be large enough to prevail in spite of relatively minor contextual conditions. To show large effects, interventions typically must introduce substantial changes in the status quo. For example, studies of school-choice interventions in which students have options to attend different schools (e.g., through vouchers, charters, or magnet schools) often show little improvement in student achievement. This small impact is likely due, at least in part, to the fact that students' new schools are not different enough from their old schools on important educational dimensions like curricula, teacher qualifications and backgrounds, and school hours.

Third, flexibility must be built into the intervention. Interventions do not operate uniformly across different sites. To effectively go to scale, the intervention must allow for some flexibility and adaptation. The challenge is determining how much adaptation to allow before the intervention is too different from the original model to be confident that it will still show effects. Effective interventions typically balance structure and flexibility. For example, instructional reforms in the Long Beach school district in California are supported by continuous improvement feedback loops that provide information about how and when to adapt the reforms to meet local needs. Similarly, the Knowledge Is Power Program (KIPP) Academies, which provide charter middle schools, use a set of standard operating principles known as the Five Pillars, but let individual principals decide how to implement them in their schools.

Mike Smith

Q Once you have determined an intervention is ready to go to scale, what challenges are involved in bringing it to new settings?

A Interventions encounter several challenges when they go to scale. For example, an intervention will face greater resistance in settings where it is seen as "disruptive" to the existing system or when it is replacing an existing intervention. In such cases, more effort may be needed to convince those involved that the new intervention is worth adopting. When done well, these disruptive interventions can cause powerful changes that lead to sizable improvements.

Related to this, the intervention's salience, duration, and intensity are especially important when replacing an existing intervention. To make the case for the new intervention, it needs to be seen as a significant improvement over the old approach. Therefore, the new intervention needs to show large and stable effects, which generally requires that it be conducted over a long period of time.

Also, bringing in a new intervention often requires relearning effective practices developed under the old intervention. For example, teachers become more comfortable with and better at teaching a curriculum as time goes on. Without even thinking about it, they do continuous improvement—they build on what they did and learned the previous year. With a new curriculum, they have to relearn the whole system and develop a new set of habits.

Finally, as I said earlier, implementation may look and behave quite differently from one place to another. If the model is sufficiently flexible, this variation should not be problematic. However, these contextual differences may mean that it takes more time and effort to see expected results.

Q What evaluation strategies should accompany the process of going to scale?

A First, evaluations of the original intervention are needed to ensure that it has internal validity—that the intervention's causal drivers are working as intended. This process involves collecting implementation data to determine that the program is implemented as planned, as well as outcome data to ensure that the program achieves what it set out to achieve. If you have that, then you know if the intervention does not work in a new setting, it is likely due to contextual differences that affect implementation rather than a flaw in the model itself.

Second, once the original site has been evaluated for internal validity and it goes to scale in new locations, evaluations are needed in the new sites to help determine what tweaks are necessary to account for differences in context. As I mentioned earlier, evaluations at this stage should emphasize ongoing learning and continuous improvement to ensure the scaling process is as successful as it can be.

Heather B. Weiss, Ed.D.
Director, Harvard Family Research Project
Email: heather_weiss@harvard.edu

Helen Janc Malone
Graduate Research Assistant, Harvard Family Research Project
Email: hjanc@mac.com

Related Resource

Smith, M. S., & Smith, M. L. (2009). **Research in the policy process.** In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 372–397). New York: Routledge Publishers (for the American Educational Research Association).

This article explores the relationship between education research and policy. In response to the impression that many education policies and interventions have little impact on education outcomes, the authors examine ways to improve the quality and usefulness of education research. The article concludes with recommendations for policymakers and policy researchers, as well as suggestions for innovations to help produce major improvements in student achievement outcomes.

Developmental Stages for Evaluating Scale

Sarah-Kathryn McDonald of the University of Chicago describes a conceptual model designed to demonstrate the role of evaluation in the scale-up process.

Promising education interventions typically go through a developmental process that starts with implementation at a single site, and then works incrementally toward the goal of scaling up their successful practices to additional settings to impact larger, more diverse populations. Evaluation can help an intervention proceed from one stage to the next, with different evaluation strategies accompanying each step of the intervention's development. The Data Research and Development Center, a research and technical center funded by the National Science Foundation as part of the U.S. Interagency Education Research Initiative, has developed a conceptual model that specifies the evaluation methods appropriate at different stages in the intervention's development. The model describes five evaluation stages that culminate in an intervention's widespread dissemination and adoption. Ideally, an intervention should proceed through each stage in order.

Stage 1: Proof of Concept. Stage 1 involves determining whether an intervention is sufficiently promising to develop and scale (including making any necessary improvements to the intervention). The goal is to produce evaluation data that can demonstrate which parts of the intervention can accommodate flexibility and which parts are not negotiable.

Stage 2: Establish Efficacy. This stage determines whether the intervention can achieve its intended results under ideal circumstances. At this stage, it is crucial that the intervention be implemented (and evaluated) with the features and in the context that are seen as optimal for success.

Stage 3: Demonstrate Effectiveness. Stage 3 aims to assess whether an intervention will achieve its objectives outside the ideal context measured in Stage 2. The objective is to establish whether an intervention "works" in a real-world setting, with all of its complications.

Stage 4: Scale-Up and Test Again. Stage 4 aims to demonstrate the intervention's impact when it is implemented among larger numbers of individuals across many contexts. This stage also examines the contextual factors that may influence the intervention's impact in different settings. In addition to identifying reasons for any observed discrepancies, these data can provide feedback to help refine the intervention or develop guidelines to ensure it operates as intended in particular contexts.

Stage 5: Postmarketing Research. The final stage explores the following questions: (a) If an intervention demonstrated to work at scale is then more widely adopted, what else do we need to learn about its effectiveness in additional contexts at larger scales? and (b) What can and should we seek to learn about the sustainability of its impact and relevance once a market has been saturated and the intervention becomes the status quo? Postmarketing research can enrich our understanding of how the impact varies among slightly different populations. It can also help identify the circumstances under which it might make more sense to adapt the intervention to the local context rather than adopt the original elements of the intervention.

Throughout the scale-up process, the context of the intervention, and both the original intervention and the new sites into which it is scaled, must be kept front and center. Because every setting has its own needs and circumstances, scale-up should not uncritically replicate the intervention. Rather, scale-up should be a context-based approach that is supported by evaluation at each stage along the way.

Sarah-Kathryn McDonald, Ph.D., MBA
Executive Director
Center for Advancing Research and Communication in Science, Technology, Engineering, and Mathematics
National Opinion Research Center at the University of Chicago
Email: mcdonald-sarah@norc.uchicago.edu

The author would like to thank Barbara Schneider (principal investigator of the Data Research and Development Center and the Center for Advancing Research and Communication in Science, Technology, Engineering, and Mathematics) for her many contributions to, and continued support for, this work. This material is based upon work supported by the National Science Foundation under Grants No. 0129365 and 0815295. Any opinions, findings, conclusions, or recommendations expressed are those of the author and do not necessarily reflect the views of the National Science Foundation.

Related Resources

McDonald, S. (2009). **Scale-up as a framework for intervention, program, and policy evaluation research.** In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 191–208). New York: Routledge Publishers (for the American Educational Research Association).

This article discusses the five stages of evaluation that educational researchers commonly use as they design, conduct, and implement evaluation studies: proof of concept, establish efficacy, demonstrate effectiveness, scale-up and test again, and postmarketing research. The article emphasizes the value of considering the primary audiences for the evaluation (i.e., who is interested in the results and who can benefit from them) as well as what these audiences plan to do with the results and in what time frame. The article also suggests how evaluators can disseminate and communicate their findings in ways that are likely to increase the chances that the data they collect will enhance policy and practice.

McDonald, S. K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). **Scaling-up exemplary interventions.** *Educational Researcher, 35*(3), 15–24.

This article has three objectives. First, it articulates the goal of scale-up research in the field of education. Second, it discusses the importance of context in conducting scale-up research. Third, it offers practical guidelines for study designs that can help education researchers provide evidence that the impacts measured can be applied to additional settings.

Early Evaluation to Inform Expansion of a Teen Pregnancy Prevention Program

Roblyn Anderson Brigham and Jennifer Nahas of Brigham Nahas Research Associates discuss the implications of their evaluation of the Children's Aid Society/Carrera Integrated School Model for expansion of the model to new school settings.

For the past 25 years, Dr. Michael Carrera and his team from the Children's Aid Society's (CAS) Adolescent Sexuality and Pregnancy Prevention Program have established a model for after school programs working with youth in middle and high school. The model relies on a holistic, long-term approach to pregnancy prevention and sexuality education that helps youth develop personal goals and the desire for a productive future. With a designation by the Coalition for Evidence-Based Policy as a "Top Tier" program based on a 3-year random assignment evaluation, CAS/Carrera has a demonstrated track record of success.[1] CAS/Carrera is now engaged in an effort to incorporate the model into the school-day curriculum at public and charter schools in targeted locations nationwide. This approach, called the Integrated School Model, began in 2006 with one school and is now operating in ten schools.

The Integrated School Model provides the seven foundational components of the CAS/Carrera after school model during the school day: daily education; weekly employment; weekly mental health classes and daily services; weekly family life and sexuality education; self-expression; lifetime individualized sports; and no-cost, comprehensive medical and dental services. Unlike other in-school programs having similar goals, the model does not simply provide additional social work staff or serve only as a health center for students. Rather, CAS/Carrera brings staff, resources, training, curricula, and expertise into partner schools and works with teachers to provide the support students in low-income neighborhoods need to succeed in school and life.

Critical Issues for Expanding the Model

To understand early implementation of the model, CAS/Carrera commissioned Brigham Nahas Research Associates to conduct a study to: (a) document how the model is implemented and integrated into the school setting during the early phase; (b) understand implementation strengths, challenges, and opportunities; and (c) document perceptions of early outcomes. The study involved site visits and in-depth interviews with the principals at each of four schools, staff of the Regional Implementation Centers, and CAS/Carrera staff working in each of the schools.

Based on this evaluation, there is reason to be excited about the possibility for expanding the model to other schools and communities. Findings from the evaluation suggest several critical issues—strengths, challenges, and lessons learned from the Integrated School Model pioneers—that can inform CAS/Carrera's thinking as the organization moves forward:

Matching philosophies. Philosophical underpinning of the model is clearly articulated, consistent, and non-negotiable. The better the match between the CAS/Carrera philosophy and that of the school, the quicker and easier it is to integrate the model into the school.

Supportive and flexible school leaders. Ease of implementation depends on the unwavering support of the school leader, a flexible school management structure, and the ability to extend school hours. It is important to look for these features in the schools that adopt the model.

Needs of students. The model fills significant and widely identified service gaps in each of the schools. Principals see CAS/Carrera as a key to addressing challenges for which they have no other solution and no other resources available. None of the schools had comprehensive mental health supports for students or a sex education curriculum in place before the arrival of CAS/Carrera.

Flexible implementation. The content of the Integrated School Model is non-negotiable, but flexible implementation of the seven core components shows that the model can be integrated into many different kinds of settings (public schools, charter schools, etc.), as long as leadership is willing to work with CAS/Carrera on operational issues such as scheduling and hiring.

continued on page 22

1. Coalition for Evidence-Based Policy. (2009). *Two new social program models identified as meeting "top tier" evidence of effectiveness.* Washington, DC: Author.

Related Resource

CAS/Carrera Integrated School Model Evaluation Report

Brigham, R. A., & Nahas, J. (2008). ***Children's Aid Society/Carrera Integrated School Model: Documentation of early implementation in four schools.*** Cambridge, MA: Brigham Nahas Research Associates.

This report discusses the methods, findings, and implications of Brigham Nahas Research Associates' evaluation of the CAS/Carrera Integrated School Model. The findings include data about implementation, which informed the lessons for scale discussed in this article, as well as perceived outcomes at the student, classroom/teacher, and school levels.

At the student level, evaluators saw evidence of improved connections with adults, management of emotions, and self-expression/communication. At the classroom/teacher level, data suggest that the program is contributing to calmer classrooms and better management of student behavior. At the school level, evaluators saw evidence of the following: (a) a shift in culture toward infusing youth development into education; (b) faster interventions in crisis situations; (c) earlier identification of students' needs; (d) better parent communication and connection; and (e) a stronger sense of belonging, cohesion, and school spirit in the cohorts served by the Integrated School Model.

For more information on the CAS Adolescent Sexuality and Pregnancy Prevention Program, visit: www.stopteenpregnancy.com.

Spreading Our WINGS: Using Performance Data to Prepare for Scale-Up

Many nonprofits collect data to measure their impact. But as Ginny Deerin, CEO of WINGS for kids, describes, they can also mine a treasure trove of performance data to improve their program models even before they undergo the scaling process.

The WINGS program began in 1996 and now serves three public elementary schools in impoverished Charleston, South Carolina, neighborhoods. WINGS is based on the premise that kids growing up in poverty often miss out on important life lessons—how to behave well, make good decisions, and build healthy relationships. WINGS instills those missing life lessons by weaving a comprehensive social and emotional education into a fresh and fun program that is delivered after school. Kids get the life lessons they need to succeed and thrive, and they get a safe place to call home after school.

To prepare for future scale-up, we mapped out a 5-year strategic plan. WINGS is now in the second year of an intensive research and development phase that requires building our organizational capacity, conducting a formative evaluation to assess our model, and conducting a summative evaluation to assess our impact.

At the heart of our evaluation efforts is a performance management system called Efforts to Outcomes (ETO). It tracks the progress of individual students toward the educational outcomes in our theory of change (TOC). Our TOC states that kids who receive a comprehensive social and emotional education for 15 hours a week within an after school program for 2 years or more will develop strong social and emotional skills and then use those skills in everyday life. It also states that they will demonstrate a strong attachment to school and consequently will graduate from high school, stay out of jail, and avoid teenage parenthood. In the long run, they will have a better chance for success and happiness in life.

The following are some of the short-term outcomes in our model that serve as stepping stones to long-term outcomes:

- Students demonstrate awareness of social and emotional skills. *Indicator: 90% can recite the WINGS Creed, which reinforces the skills in kid-friendly language.*
- Students understand the relevance of social and emotional teachings. *Indicators: 80% can cite real-life examples of applying the Creed; 80% pass testing of the social and emotional skills learning objective for the week.*
- Students demonstrate high attachment to school. *Indicator: 95% receive average or higher grades for school attendance.*
- Students use social and emotional skills in everyday life. *Indicator: 85% receive positive social development grades on report cards.*

ETO tracks each student's progress toward the outcomes and allows WINGS staff to pinpoint difficulties in mastering the curriculum's 30 learning objectives. Staff members known as WINGSLeaders carry notebooks around in which they jot down observations for each of the 10 to 12 students they supervise. They add these notations to computerized records daily.

Once we started using ETO, the data began showing us where we could make improvements in our model to better meet the needs of students. For example, WINGS staff learned that some students were not receiving positive social development grades from classroom teachers, even though our short-term outcome called for 85% to reach that target. In response, each WINGSLeader was required to develop a plan for the kids who were falling short and to share the plan with teachers in order to coordinate efforts to work with those kids.

The ETO data not only showed us deficits but also demonstrated whether the adjustments we made in response actually solved the problem. The graphic below gives an example of how ETO provides a snapshot of grades over a period of time, thus allowing us to check whether the next round of report cards indicates that our adjustments have made a difference.

ETO PROGRESS REPORT

WINGS at Memminger: SOARING in School

	As of 10/25/09	*As of 2/1/10*	*Change (%)*
Number of Kids	126	116	
SOARING *School behavior grade ≥85%*	58	73	25.86
Are Not SOARING *School behavior grade ≤85%*	68	43	36.76
No. of Staff Working Towards Results	11		
Total Time Spent by Staff	18 Hours (1,080 minutes)		

By engaging in ongoing data review and assessment, WINGS has become far more responsive and accountable. Monitoring the data at frequent intervals means that WINGS does not wait until the end of the year, or even the end of a school term, to review how staff efforts are progressing. Most important, the continuous flow and analysis of information allow us to refine and develop our program model as we prepare to scale up.

Implementing a performance management system has yielded many benefits. Although we have been results-driven since our beginning, we have sharpened our focus through the process of clarifying what factors have been driving our educational outcomes. Detailed, easy-to-access records have allowed managers to monitor whether we have been implementing the program with fidelity to our model. Real-time feedback has provided powerful motivation for our staff in aligning efforts with outcomes. But by far the most significant gain is that this continuous flow of data provides us with valuable lessons in how to improve our program model to best meet student needs for maximum effectiveness.

Ginny Deerin
Founder and CEO
WINGS for kids
Email: ginny@wingsforkids.org

Applying a Broader Concept of Scale to Evaluate a Funding Strategy

Erin Harris and Priscilla Little of Harvard Family Research Project (HFRP) discuss how HFRP used a multidimensional concept of scale to evaluate The Atlantic Philanthropies' Integrated Learning Cluster strategy.

The Atlantic Philanthropies developed its Integrated Learning Cluster (ILC) strategy as a multiyear national effort to keep disadvantaged youth ages 8 to 16 engaged in learning during out-of-school time (OST) hours. Recognizing that schools alone cannot meet the learning needs of our children, the strategy was based on the belief that extra support provided beyond the school day helps to strengthen academic skills and increase emotional intelligence. To achieve these goals, Atlantic invested in a range of nonschool supports—grants for direct service, infrastructure, and advocacy efforts.

At the heart of the ILC strategy was the need to bring the OST arena to scale to better serve disadvantaged children. But what does "scale" really mean? Traditional definitions focus on expanding the number of people served through replication or adaptation of existing models and reform efforts. The ILC strategy, however, applied a notion of scale that moves beyond this definition to encompass a broader understanding, as defined by Cynthia Coburn.[1] Her operationalization of scale has four components: spread, depth, sustainability, and shift in reform ownership.

As part of the evaluation of Atlantic's ILC strategy, Harvard Family Research Project used Coburn's definition to help identify strategic opportunities in Atlantic's investments that could enhance these four components of scale, thereby achieving the goals of the ILC strategy. For example, by applying this scaling framing to the ILC grantees, we discovered that advocacy is a key component in the scaling process and was used by most Atlantic grantees whether or not they were funded to do so. As a result, HFRP recommended that Atlantic consider incorporating advocacy investments across all of their funding strands (rather than only a subset). We discuss the components and their application to the evaluation of the ILC strategy below.

Spread is the central notion of scale as traditionally defined—the idea that an effort will expand to increase the number of programs and people served. In Coburn's definition, spread extends beyond merely "upping the numbers" to include the spread of ideas, beliefs, values, and principles that support the effort being brought to scale. Atlantic invested not only in organizations that can expand their reach, but also in resources that help introduce evidence-based, best-practice information into the national dialogue in order to build public and policy support for OST programs.

Depth refers to the nature and quality of change. According to Coburn, to be "at scale," reform efforts must effect deep and consequential change at the practice level.[2] Atlantic's investment strategy devoted significant resources to OST program models that had already achieved some "early wins" in their efforts to impact practice. Further, scaling *quality* programs (i.e., those that have undergone rigorous evaluation, have prevalent quality standards, use strategic business plans to guide their growth, and are committed to reducing barriers to program participation) was a critical part of Atlantic's overall strategy.

Sustainability refers to the idea that scale has meaning over time. As Coburn suggests, "the distribution and adoption of an innovation are only significant if its use can be sustained."[3] Atlantic's ILC investments, particularly in advocacy funding, were intended to promote and sustain high-quality OST experiences for disadvantaged youth by building awareness of and support for policies that increase federal and state funding for effective OST programs. Atlantic's early investments in organizational development and capacity building increased the likelihood that ILC grantees would continue to be active participants in the OST infrastructure long after Atlantic's investments had ceased. Further, most ILC grantees had multiple sources of private and corporate sponsorship, enabling them to leverage Atlantic's resources to garner additional support.

Shift in reform ownership, as Coburn notes, refers to shifting authority and knowledge from external actors to those who are either implementing reform or building the capacity to do so. In the case of ILC investments, these external actors were national-level OST advocacy organizations that pushed for policies to improve OST program access and quality. To ensure the success of these policies, all OST stakeholders must be on board with and take ownership of these efforts at the ground level. As noted above, many ILC grantees demonstrated this ownership by implementing policy advocacy strategies and activities whether or not they were funded to do so. This shift came about as grantees realized that advocacy efforts were necessary to achieve their goals and that they needed to become active advocates themselves rather than rely solely on others to advocate for robust OST policies. If this component of scale continues to progress among ILC grantees, then over time more grantees will internalize advocacy functions as "part of doing business," and the national advocacy organizations will build the capacity of other local-level OST programs to follow suit.

Coburn's broadened vision of scale allowed for a deeper, more nuanced understanding of Atlantic's ILC strategy investments. Although the framework was initially developed for education reform, the concepts have applications across a range of fields and types of efforts.

Erin Harris
Project Manager, Harvard Family Research Project
Email: erin_harris@harvard.edu

Priscilla Little
Associate Director, Harvard Family Research Project
Email: priscilla_little@harvard.edu

1. Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12.
2. Ibid., p. 4.
3. Ibid., p. 6.

The RALLY Program: Scaling an Inclusive Approach to Intervention and Prevention

Helen Janc Malone of Harvard Family Research Project describes an after school program's strategy for scaling its services and the role of evaluation in the scaling process.

The Responsive Advocacy for Life and Learning in Youth (RALLY) after school program takes an unconventional approach to scaling. According to RALLY's director, Gil Noam, the after school field often assumes that scaling means replicating programs. What can get lost in that approach, however, is the importance of local context, charismatic leaders, motivated staff, a mission targeted to local needs, family involvement, and youth engagement in making a program a success. Noam argues that rather than scaling up a cookie-cutter model, after school programs should invest in scaling up a model that can be customized to address the needs of local youth. As Noam explains, "People do not want another Starbucks—they want the old mom-and-pop shops."

Core Principles and Customization

RALLY, started at a middle school in Jamaica Plain, Massachusetts, has scaled its approach in demonstration sites at three more middle schools, in Honolulu, Hawaii; Syracuse, New York; and Tacoma, Washington. All of the sites apply RALLY's core principles—they offer school-based integrated services to encourage learning, promote social-emotional development, foster trusting and caring youth–adult relationships, and build youth resilience and interpersonal skills. Noam recalls communicating the RALLY approach to the individual sites: "These are the practices that we have found to be helpful and useful at RALLY. If you apply these practices to the staff you have, then you can be successful too." At the same time, each site offers its own variety of activities customized to fit each school's needs, including classroom support, tutoring, mentoring, enrichment opportunities, referrals, and professional crisis management.

The success of RALLY's approach to scaling relies in part on the selection of the sites themselves. The after school groups and school districts, with which RALLY partners, must have in place the internal capacity to align the RALLY program's approaches with their organizational missions. RALLY thus requires each site to have, at a minimum, a full-time on-site coordinator, a mental health professional, trained social workers and counselors, a strong program–school partnership, and relationships with community support services. This requisite capacity, coupled with content and organizational flexibility, allows each site to create its own strategic and business plan that considers contextual factors, adheres to the expectations of the local school district, and links to locally desired youth outcomes. As Noam explains, RALLY's approach to scaling has allowed each of the sites to operate as a cohort, exchanging ideas that work for that community.

Evaluation Strategy

The RALLY scaling method has important implications for its evaluation strategy. Flexibility in local implementation allows for a customized evaluation at each site. According to Noam, many after school programs that operate in multiple sites engage in evaluations that focus mainly on accountability (i.e., to satisfy a funder). These evaluations rely heavily on standardized measures that do not adequately assess individual programs' desired outcomes; their reporting time often lags months behind when the data were collected, diminishing the ability of site staff to make continuous, informed decisions about activities and services. A more beneficial approach, Noam suggests, is to focus on data-driven programming where the evaluation is directly connected to youth experiences at the program sites. Such evaluation strategies create a feedback loop around youth experiences, staff expectations, and program goals. Feedback from frequent participants in programs helps guide management decisions, aids in the construction of activities that engage youth in learning, and improves program services and supports.

Scaling up programs across multiple sites is a popular practice in the after school field, explains Noam, but "letting all flowers bloom" requires investing in a strong organizational infrastructure, developing strategic plans on how to approach scale, and engaging in rigorous evaluation that offers data on the quality of implementation across sites. For programs with the capacity to scale up to multiple sites, employing a flexible program model that encourages local adaptation in services, activities, and evaluation strategies can create a sense of ownership at the sites. It can also create a cohort model of learning that encourages individual site improvement and growth and, potentially, an increased quality of the program as a whole.

Helen Janc Malone
Graduate Research Assistant, Harvard Family Research Project
Email: hjanc@mac.com

Related Resources

Program in Education, Afterschool & Resiliency (PEAR). Dedicated to a whole-child approach, PEAR integrates research, theory, and practice across youth development, school reform, and mental health fields. Gil Noam, PEAR's director, is also the editor-in-chief of *New Directions for Youth Development,* a quarterly journal dedicated to contemporary issues in the field of youth development. To learn more, visit: www.pearweb.org. More information about the RALLY program can be found at: www.pearweb.org/rally.

Assessment Tools in Informal Science (ATIS). PEAR's new database, ATIS, offers assessment tools for after school programs. Users can search for tools by age, domain, type, or custom criteria. Each tool provides users with a short description, psychometric properties, and an author contact. Available instruments can quickly be applied in after school settings to start program- and youth-level data collection. To access the database, visit: www.pearweb.org/atis/.

White House Office of Social Innovation and Civic Participation

Harvard Family Research Project describes the functions of the White House Office of Social Innovation and Civic Participation and its oversight of the federal Social Innovation Fund.

In response to the perceived lack of innovation and use of rigorous approaches to identifying "what works" in the nonprofit sector, last year President Obama created the Office of Social Innovation and Civic Participation. The office is a part of the Domestic Policy Council, the entity that coordinates domestic policy making in the White House.

In April 2009, Sonal Shah, who previously headed Google's global development team, the search engine's philanthropic arm, was appointed to lead the new office. Prior to Google, she had worked for think tanks including the Center for Global Development and the Center for American Progress as well as for the U.S. government in the Department of Treasury.

The new office seeks to address the need to identify and scale up successful nonprofit initiatives and has four primary goals. First, it aims to develop partnerships between the government and nonprofits, businesses, and philanthropists. Second, it works to support the rigorous evaluation and scaling of innovative, promising practices. Third, it supports the use of new media tools to encourage greater civic participation. Lastly, it promotes national service.

The office emphasizes the value of government funding for bottom-up initiatives—that is, nonprofit initiatives that are driven by those outside of government. Speaking at the 2009 Global Philanthropy Forum Conference, Shah told donors and grantmakers, "Partnerships matter a lot to us. It's not just that the government does, and everybody else follows. It really is about where can we learn and what can we do differently."[1]

The office welcomes ideas and innovations from new and long-standing organizations alike, as Melody Barnes, director of the Domestic Policy Council, stressed in an interview with the *Chronicle of Philanthropy:* "Often when there's change, people believe we are leaving what's been working over time behind. The idea is to add capacity and to bring new ideas and new people and new models to the fore and to the task of addressing these big challenges. This shouldn't be seen as the White House turning its back on the more traditional allies and moving on to something new. An organization that's been around for 80 years can have an innovative idea, an idea that will have significant impact and build capacity."[2]

Social Innovation Fund

The Social Innovation Fund is a priority for the new office. The office is working in collaboration with the Corporation for National and Community Service, the federal agency that runs national service programs, to set up the Social Innovation Fund pilot program. This program was authorized in the Edward M. Kennedy Serve America Act of 2009, which expanded the national service programs administered by the Corporation for National and Community Service. According to Nicola Goren, acting CEO of the corporation, "The goal is to build a pipeline of organizations and practices with strong evidence, and the capacity to grow and increase the impact of their work. The Social Innovation Fund will provide the support needed to help move organizations from the promising stage to the stage where they have more concrete evidence that what they do works."[3]

In the FY2010 budget, President Obama requested $50 million for the Social Innovation Fund program. The draft Notice of Funding Availability for this fund was made available to the public for review in December 2009. According to the press release for the Fund, every dollar in federal funding must be matched by three dollars in private funds. The plan is to provide 5-year grants, ranging from $5 million to $10 million, to between five and seven intermediaries (i.e., grantmaking organizations) in FY2010. These intermediaries will then provide subgrants to community organizations to identify and scale promising, results-oriented nonprofit programs. In addition, the intermediaries that receive federal funding are required to provide other supports (besides funding) to the identified promising programs, including assistance with management, fundraising, and evaluation needs.

Harvard Family Research Project

1. Wilhelm, I. (2009). White House Social Innovation office to have three goals. *The Chronicle of Philanthropy.* Retrieved January 14, 2010, from: http://philanthropy.com/news/conference/7957/white-house-social-innovation-office-to-have-three-goals.
2. Perry, S. (2009). White House officials discuss plans for Social-Innovation Office. *The Chronicle of Philanthropy.* Retrieved January 14, 2010, from: http://philanthropy.com/news/updates/8381/white-house-officials-discuss-plans-for-social-innovation-office.
3. Goren, N. (October 8, 2009). Replicating success. *White House Blog*. Retrieved January 14, 2010, from: www.whitehouse.gov/blog/Replicating-Success.

Related Resources

The **press release about the creation of the White House Office of Social Innovation and Civic Participation** can be found online at: www.whitehouse.gov/the press office/President-Obama-to-Request-50-Million-to-Identify-and-Expand-Effective-Innovative-Non-Profits.

Information on the **Social Innovation Fund** can be found on the Corporation for National and Community Service's website at: www.nationalservice.gov/about/serveamerica/innovation.asp. This website includes press releases, official statements, and updates on the Fund, including information about the status of the Notice of Funding Availability.

Serve.gov is an online resource managed by the Corporation for National and Community Service to help citizens find volunteer opportunities and to facilitate the creation of new ones. It is part of United We Serve, a national service initiative. For more information, visit: www.serve.gov. On this website, you can create your own project, recruit volunteers, find volunteer opportunities, and read and share stories of people's experiences in volunteer service.

New Releases on Benchmarking Electronic Communications

Katie Chun of Harvard Family Research Project describes new studies that help nonprofits assess the success of their online and text-messaging strategies.

In 2007, *The Evaluation Exchange* featured the *eNonprofit Benchmarks Study,* a research effort by M+R Strategic Services and the Nonprofit Technology Network that developed metrics to measure the effectiveness of nonprofit online advocacy and fundraising efforts.[1] Recent releases from this effort and others offer new findings on benchmarks for electronic communications.

2010 Nonprofit Social Media Benchmarks Study

This study by M+R Strategic Services helps nonprofits develop meaningful metrics for their social media efforts. It focuses on Facebook and Twitter as the most popular social media platforms. The Facebook analysis looked at the activity of five well-known nonprofits.[2] The Twitter analysis examined data on 10 diverse organizations that are among the most successful based on their number of followers.[3]

Highlights include:

- Organizations posted on their Facebook pages an average of 6 times per week; they "tweeted" 4 or 5 times a day.
- An average of 2.5% of an organization's Facebook fans responded to posts ("liked" or "commented") each week.
- Facebook metrics can include fan-base size and growth, fan churn (number who come and go), posts by organizations, page views, and fan actions.
- Twitter metrics can include organization tweets, number of followers, and follower churn.

2010 Nonprofit Text Messaging Benchmarks

This first-of-its-kind report from M+R Strategic Services and MobileActive.org offers benchmarks for text messaging and describes how nonprofits are using this medium. The study looked at list size, broadcasts, fundraising, and advocacy data from six organizations.[4] Key findings include:

- More than 80% of text subscribers joined by entering their mobile phone number on sign-up pages on a nonprofit's website.
- Organizations studied sent an average of 1.6 text messages per month.
- The response rate for advocacy text messages that asked recipients to promote an issue with key decision makers was 4.7%, nearly six times the response rate for emails that make similar requests.

2009 eNonprofit Benchmarks Study

This study from M+R Strategic Services and the Nonprofit Technology Network updated findings from their original study on online messaging, fundraising, and advocacy. Data reported in the new study are from 2007–08 and were submitted by 32 leading nonprofit organizations (the original study used 2003–05 data).

The 2009 study responded to questions about, for example, the effects of the financial crisis on online fundraising and email response rates. Data were collected from key national nonprofits and major providers of nonprofit online communications and marketing programs, as well as from an online survey of the broader nonprofit community.

The new data show that the 2 years before the study's release were a time of exponential growth and innovation for online activism. Many nonprofits, for example, tried to capitalize on the success of popular social networking websites like Facebook and Twitter. Many also tried to emulate the online strategies used by the 2008 presidential campaigns.

Other highlights from the study include:

- The rate at which recipients opened fundraising, advocacy, or newsletter emails and clicked on links within them decreased slightly from 2007 to 2008. However, email response rates (i.e., the rate at which people completed an action requested by an email, such as making a donation) did not drop significantly; the authors suggest that, over time, email subscribers become familiar with advocacy and donation landing pages and do not feel the need or have the curiosity to click on a link before responding to the message in the email.
- Email lists continue to be an effective way for nonprofits to reach supporters. Despite the growing popularity of social networking sites that compete with email (35% of U.S. adults now have an account with one or more social networking websites[5]), response rates for fundraising and advocacy emails remained relatively steady from 2007 to 2008.
- The number of online gifts to nonprofits increased 43% in 2008 over 2007, while the total monetary amount increased by only 26%. Even though more people donated to nonprofits in recent years, the average gift size in 2008 was $71, which is $15 less than in the previous year.

To learn more about these studies and to access new reports as they are released, go to: www.e-benchmarksstudy.com.

Katie Chun
Graduate Research Assistant, Harvard Family Research Project
Email: katie.chun@post.harvard.edu

1. Matheson, K. (2007). The eNonprofits benchmarks study: Diving into email metrics. *Evaluation Exchange, 8*(1), 25. Retrieved July 20, 2009, from: www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/advocacy-and-policy-change/the-enonprofit-benchmarks-study-diving-into-email-metrics.
2. Human Rights Campaign, Oxfam America, NARAL Pro-Choice America, Easter Seals, and the League of Conservation Voters.
3. charity: water, CARE, Acumen Fund, Live Earth, Red Cross, Greenpeace, Brainline.org, Oxfam, Human Rights Campaign, and NARAL Pro-Choice America.
4. American Society for the Prevention of Cruelty to Animals, Planned Parenthood Federation of America, NARAL Pro-Choice America, The Humane Society of the United States, Defenders of Wildlife, and Human Rights Campaign.
5. Lenhart, A. (2009). *Adults and social network websites.* Washington, DC: Pew Internet & American Life Project. Retrieved July 23, 2009 from: www.pewinternet.org/Reports/2009/Adults-and-Social-Network-Websites.aspx.

Why Facebook Matters for Nonprofits

Katie Chun of Harvard Family Research Project discusses the growing momentum and collateral challenges of Facebook as the next major vehicle for nonprofits.

What Is Facebook?

Facebook is a social networking site—a website that allows users to join a virtual community and communicate with others who have common interests. These websites represent a paradigm shift in how we use the Internet. Rather than serving as a one-way resource for obtaining information, social networking sites are "relational hubs" that facilitate a wide range of online interactions: establishing friendships, maintaining long-distance relationships, strengthening professional connections, mobilizing support for causes, and others. Social networking sites enable users to be active participants in their Internet experience.

Developed in 2004, Facebook is the largest social networking site on the Internet, with more than 250 million active users worldwide[1] and a global audience that is growing at an annual rate of 153%.[2] Facebook is free to use, and anyone with a valid email address can join. After individuals register for a personal account, they create a profile that contains personal and professional content.

Facebook users then connect with other users by adding them as "friends." Only logged-in Facebook friends can view each other's full profiles, and friends are notified of each other's Facebook activity. Friends can find each other by using Facebook's search engine or by browsing through "networks," groups of users connected around a common interest, school, or workplace.

Facebook offers a host of privacy preferences, letting users choose what information is available to whom. These privacy settings are one reason for Facebook's enormous success compared with other social networking sites.

Why Facebook for Nonprofits?

Facebook has emerged as a user-friendly site for nonprofits, largely because its "Pages" feature gives them a way to create a quick and cost-free presence on the Internet. A Page is similar to a profile, except that a Page represents an entity—a nonprofit, a local business, a brand or product, or a public figure—rather than an individual person. Facebook users can "become a fan" of a Page, thereby connecting themselves with that entity and showing public support for it. Pages are useful because they facilitate interaction between the larger entity and the public. Nonprofits may find Facebook especially helpful in accomplishing the following activities:

Fundraising. Organizations can conduct fundraising through Facebook using a feature that allows them to receive monthly payments of funds donated by Facebook members. Facebook provides a virtual "scorecard" that allows the public to view fundraising progress, and individuals can post the causes they support on their personal profiles. The goal is to achieve a snowball promotion effect, fostering a virtual community of Facebook friends with similar interests that align with the missions of various nonprofits.

Information dissemination. Facebook Pages include functions that allow organizations to promote their work by posting mission statements, news, contact information, details on upcoming events, and other items related to their work. Quick and widespread dissemination of information to the Facebook community is useful when spreading the word about events and activities and for increasing attendance at conferences, workshops, and volunteer opportunities. In addition, Facebook's visibility and global accessibility enables nonprofits to increase the number of downloads or purchases of a product or publication. Further, Facebook saves nonprofits money on postage and mailing expenditures by allowing them to conduct mass communication online.

Facebook provides a virtual pathway for the public and key stakeholders to connect personally with nonprofits.

Mobilizing key stakeholders and the public. Facebook can serve as a platform for rallying public support. Virtual mobilization can spur large numbers of people to write government officials, corporate representatives, or the media about the nonprofit's cause. Facebook provides a virtual pathway for the public and key stakeholders to connect personally with nonprofits, creating a sense of openness and transparency. In addition, the social nature of Facebook allows nonprofits to interact with potential supporters relationally rather than pushing the fundraising agenda alone. For example, Facebook users can actively participate on a Page by posting messages and engaging in discussion boards. According to one nonprofit social media expert, Facebook's fundraising feature "isn't just about raising money, it's also about raising friends and awareness, and in the long run, turning loose social ties into stronger ones for a cause may be more important than one-time donations of $10 and $20 right now."[3] Another nonprofit social media expert notes that the fundraising function on Facebook is "a friending tool, not a fundraising tool."[4]

Data collection. Facebook offers nonprofits both conventional and creative methods to collect data for internal tracking and accountability and to inform decision making. Ways in which Facebook can function as a tool for data collection include:

- *Online surveys*. Nonprofits can use Facebook to advertise, recruit for, and administer surveys online. This data collection method can be used to evaluate any number of topics: how people heard about the nonprofit, why repeat donors give, what changes occur in attitudes or awareness, and so on.

1. Zuckerberg, M. (July 15, 2009). Now connecting 250 million people. Message posted to: http://blog.facebook.com/.
2. Schonfield, E. (August 12, 2008). Facebook is not only the world's largest social network, it is also the fastest growing. Message posted to: www.techcrunch.com/.
3. Fine, A. (April 22, 2009). Washington Post disses causes on Facebook. Message posted to: http://afine2.wordpress.com.
4. Kantor, B. (April 22, 2009). Hello, Washington Post: Dollars per Facebook donor is not the right metric for success. Message posted to: http://beth.typepad.com.

- *Online tracking.* Nonprofits can use Facebook to track financial donations, membership numbers (including the number of active members per week), demographics (e.g., gender and geographic location), and the types of interactions members have with the organization's Facebook Page.[5]
- *Social "listening."* Comments or questions are often posted on nonprofit-affiliated Facebook spaces, such as message boards or blogs. By using Facebook to review conversational exchanges with the public, a nonprofit can keep an ear turned toward potential volunteers, donors, and clients and learn more about who they are and what their interests are.

Limitations of Facebook and Other Social Networking Sites

Facebook and other social networking sites can help nonprofits supplement communication with their key audiences, but should not replace face-to-face interactions and other ways of engaging stakeholders. While a nonprofit's presence on Facebook can reach a wide audience, it excludes people who do not have Facebook accounts (and are not interested in getting an account).

The nature of online communities presents additional limitations and challenges in terms of translating online representation to off-line action. A direct correlation does not exist between the two.[6] Facebook Pages may be easy to join with a simple click, but there is the potential for people to feel that they are "doing good" just by clicking and associating themselves with a cause rather than involving themselves in it. The challenge lies in using Facebook as a catalyst and springboard for real-world opportunities.

Katie Chun
Graduate Research Assistant, Harvard Family Research Project
Email: katie.chun@post.harvard.edu

5. Smith, J. (May 5, 2009). Facebook upgrading "Insights" metrics dashboard for page managers tonight. Message posted to: www.insidefacebook.com.
6. Hesse, M. (July 2, 2009). Facebook's easy virtue. Retrieved July 17, 2009, from: www.washingtonpost.com/wp-dyn/content/article/2009/07/01/AR2009070103936.html.

Save the Children

continued from page 9

Mintz and her colleagues review evaluation findings with each program site to identify areas for improvement and to chart progress over time in order to achieve the largest gains for the greatest number of struggling readers. Evidence about learning needs among the program's youngest participants was a factor prompting the development of STC's emerging reader program for kindergarten and first-grade children.

The most recent annual evaluation report in the series, which is directed at Policy Studies Associates by Richard White, is available at: www.policystudies.com/studies/school/STC.html.

Elizabeth R. Reisner
Principal
Policy Studies Associates, Inc.
Email: ereisner@policystudies.com

Ask the Expert

continued from page 11

use that money to expand to more sites or to increase the number of sites they have in some of their replication communities. On the other hand, it would be very helpful to their long-term growth to conduct that experimental design impact study. It is a tough call, but they have the information they need to make that call." Thus, if Citizen Schools chooses to emphasize further replication, it could broaden its reach and serve more children, but it would be missing an opportunity to further solidify its program credibility and proof of impact.

The Iowa PIRC's experience demonstrates that deepening program service components makes meaningful impact more likely, but can also lead to a retraction of program *reach,* since the emphasis falls on a particular site (or small set of sites). Depending on program service mandates, this trade-off can have implications for how outside observers assess the efficacy of the program; administrators must make difficult decisions about whether to strive for broad reach, which might not yield much in the way of measurable impact, or to focus on meaningful service depth, which provides less "ground cover" and thus less visibility for the program as a whole.

Heidi Rosenberg, Ph.D.
Senior Research Analyst, Harvard Family Research Project
Email: heidi_rosenberg@harvard.edu

Helen Westmoreland
Director of Program Quality
Flamboyan Foundation
Email: hwestmoreland@flamboyanfoundation.org

Teen Pregnancy Prevention

continued from page 15

Regional management and support. The strong regional management structure and training provided by CAS/Carrera are critical to ensuring the integrity of the model going forward.

If CAS/Carrera continues to expand into additional schools, and if the expectations for the Integrated School Model's impact are realized and documented, education leaders—from principals to policymakers—will take notice, and a network of highly effective urban schools is likely to result. The work in these pioneer schools could transform the way many schools operate, the approaches teachers take, and the way children experience education.

Roblyn Anderson Brigham, Ph.D.
Partner
Email: roblyn@bnra.net

Jennifer Nahas, M.M.H.S.
Partner
Email: jennifer@bnra.net

Brigham Nahas Research Associates

An annotated list of papers, organizations, initiatives, and other resources related to the issue's theme.

Resources on Scaling Impact

Borman, G. D. (2009). ***National efforts to bring reform to scale in America's high-poverty elementary and secondary schools.*** Washington, DC: Center on Education Policy. This paper examines the history of reform in high-poverty elementary and secondary schools with a focus on the challenges involved in scaling such reform. Additionally, it explores the political climate surrounding school reform and notes that growing interest among policymakers could lead to investments in scaling quality programs to more of America's high-poverty schools. Online at: www.cep-dc.org/_data/n_0001/resources/live/NationalEffortsBringReformScaleAmericasHighPovertySchools.pdf.

Grossman, A., & Curran, D. F. (2003). ***Harlem Children's Zone: Driving performance with measurement and evaluation***. Boston: Harvard Business School. In this case study, the authors examine the methods used to evaluate the Harlem Children's Zone as it went to scale. The Harlem Children's Zone offers educational, medical, and social services to children from birth through college in a New York City neighborhood. It provides an important example of successful scaling and the role of evaluation in the process. Online at: http://hbr.org/product/harlem-children-s-zone-driving-performance-with-me/an/303109-PDF-ENG.

McDonald, J., Klein, E., & Riordan, M. (2009). ***Going to scale with new school designs: Reinventing high school***. New York: Teachers College Press. In this book, the authors analyze a scaling approach to redesigning American high schools (i.e., identifying successful local solutions that can be scaled and brought into new areas). The text includes commentaries on the approach and practical advice on confronting the challenges of introducing new programs into communities. In particular, it addresses such challenges as adaptation, teaching, ownership, communication, feedback, resources, and politics.

Ottoson, J. M., & Hawe, P. (Eds.). (2009). **Knowledge utilization, diffusion, implementation, transfer, and translation: Implications for evaluation.** *New Directions for Evaluation, 124*. This special issue of the journal offers a series of comprehensive theory and literature reviews on topics that are relevant to any evaluator examining the scale-up process. A chapter is devoted to each of the following subjects: knowledge utilization, innovation diffusion, policy implementation, the transfer process, and knowledge translation. Each chapter describes disciplinary roots, assumptions about change, key variables in the process, and contextual influences.

Roob, N., & Bradach, J. (2009). ***Scaling what works: Implications for philanthropists, policymakers, and nonprofit leaders.*** New York: Edna McConnell Clark Foundation and The Bridgespan Group. This paper offers recommendations for turning small, flourishing programs into large-scale successes. Examining examples of successful scaling in the for-profit and nonprofit sectors, the authors determine that increasing the scope of a program requires significant resources. They argue that the tendency of nonprofits to cut corners when they receive inadequate funding does not allow for successful scaling. Instead, they suggest a strategy that involves larger amounts of funding going to fewer sources, extensive evaluation and reporting to determine which programs to fund, and stronger partnerships between funders and the nonprofit sector. Online at: www.emcf.org/pdf/paper_scalingwhatworks.pdf.

The Social Impact Exchange is a forum for people with diverse careers and interests to share knowledge and learn about scaling effective social programs and solutions. It provides a gathering place for those looking to fund large-scale expansions of nonprofit initiatives. Online at: www.socialimpactexchange.org.

Other New Resources, Ideas, and Efforts

Cause Communications' Online Outreach Tools Guide is a quick-reference guide for groups considering how to most effectively use the Internet and interactive technologies to further their policy change agenda. Demystifying "high-tech" terms, the guide looks at Web 2.0 tools such as wikis, blogs, social networking sites, and Twitter, and assesses their relative strengths and weaknesses for achieving communications and policy objectives. The guide also suggests quantitative and qualitative metrics to use with each tool. Online at: www.hersheycause.com/pdf/online-outreach-tools-guide.pdf.

The Center for Evaluation Innovation, a new nonprofit effort based in Washington DC, is building the field of evaluation in areas that are challenging to measure and where traditional approaches are not always a good fit. The center specializes in advocacy evaluation and other emerging areas (e.g., systems change and communications evaluation) where fresh thinking and new approaches are needed. Online at: www.evaluationinnovation.org.

iScale is a social enterprise group that seeks to promote positive change by spreading innovative ideas so they have large-scale impact. iScale's activities include action learning projects, community building, and networking to share innovations. iScale addresses problems in areas ranging from finance to the environment to corruption; most projects have an international or worldwide focus. Online at: www.scalingimpact.net.

Platinum standard is a term used by Deborah Lowe Vandell, chair of the Department of Education at University of California at Irvine, to describe an evaluation design that uses a range of measures (e.g., observations, interviews, surveys, etc.) to collect qualitative and quantitative data on program implementation and outcomes. This concept is a response to the *gold standard,* an evaluation approach that uses randomized clinical trials to collect outcome data. Vandell argues that the *platinum standard* can provide a fuller picture of a given program than the *gold standard,* because it incorporates both implementation and outcome data to show how program elements relate to outcomes rather than focusing on outcomes alone. This approach is particularly useful for programs that are going to scale; it can help them determine whether they are successful, and thus worthy of going to scale, and what elements have led to that success, thus enabling them to choose which program components to scale.

Related Resources

Online Conversations about Scaling Impact

Because the idea of scaling impact is a hot topic in the nonprofit world, a number of informal conversations are occurring online about this topic (including the implications for evaluation), especially on blogs. Links to some of these recent conversations, which represent a wide range of views and opinions, can be accessed at: www.hfrp.org/ScalingResources.

Harvard Family Research Project

Harvard Graduate School of Education
3 Garden Street
Cambridge, Massachusetts 02138
hfrp_pubs@gse.harvard.edu
www.hfrp.org

nonprofit org.
US postage
PAID
Boston, MA
Permit #56376

> end notes

Scaling Impact: Six Takeaways

HFRP summarizes key observations in this issue of The Evaluation Exchange.

Scaling impact often refers to scaling programs or interventions, but ideas, technologies, skills, and policies can also be scaled. Thinking about scale more broadly can reveal additional opportunities to scale impact beyond the traditional business model of replication.

- *Scaling impact means more than just upping the numbers served.* Scale is typically conceptualized as serving more people or populations. However, scaling can also mean deepening or sustaining impact. It is not just about increasing quantity; it can be about increasing quality.
- *Local context matters; adaptation may be necessary.* Unlike scientific models of scale, social-sector scaling is subject to uncontrollable and unpredictable factors that can affect the final result. What works well in one setting might not work as well in another. For this reason, flexibility and adaptability may be necessary to achieve success when a program, intervention, or idea is translated to new settings.
- *Evaluation can play multiple roles in the scaling process.* In the past, nonprofits usually employed evaluation at the front end of the scaling process—for example, to determine whether something was worthy of going to scale. But evaluation has other important roles to play. It can identify whether an organization is ready or has sufficient capacity to take something to scale or whether potential settings are good candidates for expansion. It can help identify the most appropriate method for going to scale. And it can assess how well the scaling process is working and provide feedback on how to improve it if necessary. Lastly, nonprofits can use evaluation to demonstrate the success of the scaling process when making the case to funders and other stakeholders that efforts should be scaled up even more. Evaluation that is integrated or embedded within the scaling process can make an important contribution to its success.
- *A slow and deliberate approach is likely to be more effective than rapid scaling.* Although this point may seem obvious, it is easy to get caught up in the excitement of scaling quickly without thinking through all of the necessary steps to ensure success. Taking the time to make sure that the initial intervention, idea, or strategy works well, and testing to make sure it is likely to show similar results in new settings, can save considerable time and money in the long run. Evaluation can help ensure that all of the necessary ingredients for success are in place.
- *Scaling is a long-term process with long-term evaluation needs.* Even with careful planning, unforeseen issues will come up during the scaling process. Flexibility and a commitment to continuous monitoring and evaluation are required. Scaling is a long-term process that requires vigilance to ensure that it is working as planned and a dedication to making adjustments when it is not. Evaluation can play a crucial role in helping to identify potential problems and suggest possible solutions.

Scaling is a long-term process that requires vigilance to ensure that it is working as planned.