

Scaling up the Replication Ladder: Considerations in Evaluating Scale-up Efforts

American Evaluation Association Annual Conference

Think Tank Session #443

Thursday, November 3, 2011

Facilitators:

Ginger Fitzhugh (gfitzhugh@eraeval.org)

Carolyn Cohen (ccohen@seanet.com)

Julie Elworth (jelworth@eraeval.org)

Davis Patterson (davispatterson@comcast.net)

David Reider (david@educationdesign.biz)

This Think Tank group discussion surfaced a number of themes related to evaluations of projects that are scaling, expanding, or promoting innovations. Here are some of the big ideas, questions, or common themes, followed by more detailed notes captured during the session:

- Scaling can involve programs, ideas/innovations, technology/skill, policy (see Evaluation Exchange vol. XV, no. 1, on Scaling Impact; Harvard Family Research Project).
- How can scaling be accomplished successfully across diverse contexts, and what are standard ingredients that an evaluation should attend to? Key inputs, implementation/process, and contextual factors (not exhaustive and not mutually exclusive) suggested included
 - Environmental influences on implementation of scale-up
 - Dosage
 - Management
 - Available resources
 - Values
 - Fidelity, including monitoring fidelity to original values behind intervention
 - Multi-organizational factors: institutionalization of norms, guidelines, tools, documents, etc.
 - Stakeholders' diverse investments and values
 - Momentum/political will behind effort
- How does one identify essential elements of a model (necessary, if not sufficient) for scaling successfully?
- The scale of scaling matters. A spreading innovation—in a community, a society, or worldwide—can have positive benefits as well as unintended or potentially harmful consequences. Programs and evaluators need to think about these possibilities. When is enough? How can we monitor and evaluate the “right-sizing” of programs, innovations, etc.?
- It's important to define success and failure, and to be able to measure success well. It's also important to be able to capitalize on actual outcomes that are beneficial even if not originally intended (i.e., not “success” as originally conceptualized) and explore possible impacts and value.
- Positioning for potential scale-up requires thinking ahead from the beginning of a project. Evaluators can also play a developmental/strategic learning role with organizations as the initial program develops and as scaling occurs.
- More research/funding to study scale-up is needed.

Detailed notes

Questions and issues about scaling/scale-up of interest to session participants:

Framework for monitoring policy contribution, adding a third dimension of scale?

How do you scale up across different contexts, e.g., remote villages? Context can change rapidly even within one locale, e.g., different parts of Ethiopia.

What are key areas to focus on at different scales of scale?

Identifying implementation factors related to impacts; specifying more carefully and testing what implementation factors affect outcomes

Identifying core components across cultures, sites in implementation of scaleup—vision from above, below, etc.

We know what worked in Austin. How do we figure out what is really essential about the model and document this?

Link between evaluation and site support—the evaluation I'm working on is guiding a lot of the implementation

What's the current thinking, beyond the minutiae of implementation—fidelity, etc.—how can I think bigger about scale-up?

I work internationally; what's going on domestically? To guide strategic planning and evaluation—what frameworks are there? How distillable is our framework for reproductive health innovation—can it be simplified or made generic enough to be useful to other organizations scaling up?

How do you replicate an idea—a concept that's not fully/explicitly defined?

The program has loosely defined models. The program cares about outcomes being achieved; we (the evaluation) care about how/implementation.

What factors in the implementation environment are conducive and must be identified to successfully scale up?

Interest in policy evaluation

What are existing or possible frameworks for evaluating scale-up efforts

Association of a standard set of factors with variation in impacts; investigating application of factors to others models than Welfare-to-Work programs. Combination of standard factors across all types of human service programs: dose, management factors, fidelity, resources, environment (e.g., poverty): test across different program models that are having an impact evaluation

Read Harvard Family Research Project Evaluation Exchange issue on Scaleup

Similar frameworks from implementation evaluation to scaleup evaluation? See Grover, ...

Systems-oriented framework: two axes—(1) extending out service access, (2) institutionalization (norms, standards, supervision tools, guidelines). What are institutionalization guidelines and documents we want to change? Also want to provide evidence to policymakers that it's a good investment. Framework has a values element. Literature shows that proof of concept has inherent values in it. How do you monitor that as you scale up—so that values don't get lost?

Systems approaches are the key. Scaling up from 200 farmers to 1.2 million: scaling up technology is easy, but affects other parts of the system (collapse in price of teff). Loss of profitability leads to opting out of the system. Scale of the scale up is important to monitor: when is enough?

Challenges of looking at global best practices and suitability for replication in various parts of the world.

How do we evaluate that it's happened, happened well, not having negative effects? Numbers are easy, to say something has been replicated, but harder to determine if you should keep replicating in different contexts. How to monitor and evaluate effectiveness of scaling—is it maximizing its potential (e.g., ridership on bus rapid transit)? Is it displacing something else you don't intent to displace?

Stakeholders' investments in scale up may be very different: need to anticipate and measure 2nd and 3rd order effects/externalities: domino effects; initial idea that is successful and expands has second order effects

Transferability and commonality across different types of scale. What are "climate" factors, how fertile is the ground—need to identify the context

Federal policy to state to local level differences, rural, suburban, inner city—how do differences in context affect outcomes? Key components cut across, and local policies are distinct

Models designed for public sector v. private sector

What are key questions you would ask as an evaluator to start, beginning to work with clients to evaluate a scaling effort? (vs. other kinds of evaluation that don't involve scaling)

1. Is there political will behind the effort? Is policy framework right? Is there something to motivate it beyond implementation? Momentum? If there are barriers, that would inform next steps.
2. Look at innovation itself—what is being scaled up? Is it the first time, has it been successful elsewhere? If so, where, under what conditions, when has it failed?
3. Depends on stage: formative, retrospective. Looking at success factors and key conditions for replication across multiple countries/sites. Don't get bogged down in counterfactuals, if you want rapid catalytic change.
4. Is your outcome straightforward (e.g., binary), more complex?—need to define what it is you're trying to reproduce. There's a difference between not defining success and failure vs. having to come up with sensitive, complicated measures of success

What do you want to be thinking of at a more formative stage when you think client might be scaling?

1. Become very clear on level of potential scale. Makes a difference whether you're talking about parents of schoolchildren as stakeholders vs. national leaders (macro/micro questions). Community level interventions haven't had a history of replicating well.

Strategic planning role as an evaluator?

1. What might organizational outcomes be? Cost considerations, staffing needs, training models? Capacity questions
2. Scaling has to be multi-organizational—what do you measure as changes in organizations' capacity? Technical, elusive leadership capacities? What changes will the organization be willing to make to keep something at scale?
3. Keep an eye out for emergent/unintended outcomes—spinoffs, transformative things. Infuse language with notion of sustainability from the get-go. A process of thinking long-term. Pay attention to changing contexts rather than/in addition to being fixated on outcome. If outcomes don't show up, maybe they're not that germane to the innovation. What validity/value is there to outcomes actually observed? It can be advantageous to shift the outcomes to something the program is actually impacting, especially in a complex field/intervention. Acknowledging where and why it didn't work as intended and then exploring potential impacts with funder—could it lead to change in RFP?
4. Who is responsible for the evaluation? Looks very different in different places, depends on level of resources.
5. Scale of evaluation itself—multiple levels of evaluation (individual program level, scaling aspect). Evaluation of scaling itself—see Managing Complex Change Framework.

More research and funding to study scale up is necessary!