Lessons Learned on ‘Scaling Up’ of Projects

By Debra Viadero

Having developed a technology-based teaching unit on weather that appeared to work well for middle school students, Nancy Butler Songer and her colleagues at the University of Michigan decided in the late 1990s to take the next logical step in their research program: They scaled up.

Over two years, they recruited tens of thousands of teachers, students, and scientists from around the world to share data and take part in their Kids as Global Scientists computer network. But the researchers found it difficult—if not impossible—to reliably assess what was happening in all those schools and classrooms. Had their software program attracted a wide range of teachers or just the mavericks who were hungriest for change? Were teachers carrying out the lessons faithfully?

“We quickly realized that measuring impact by how many people does your intervention touch or influence was a pretty simplistic way of looking at it,” said Ms. Songer, a professor of science education and learning technologies at the campus in Ann Arbor, Mich. “The integrity of the program could be anything.”

That lesson is among dozens collected in a new series of books that distills the “scale-up” experiences of projects financed under the just-ended Interagency Education Research Initiative, or IERI. Arguably the largest cross-disciplinary education research effort the federal government has ever undertaken, the initiative was a collaboration of three agencies: the U.S. Department of Education, the National Science Foundation, and the National Institute of Child Health and Human Development. The effort provided multiyear grants averaging $5 million to $6 million each to 101 different projects at various stages in the research pipeline.

While not all those projects went to scale, the federal funding stream was sustained enough and deep enough so that many grantees could try to replicate their successes in a wider range of educational settings. In the process, the researchers accumulated evidence for the rest of the field on what works best, for whom, and under what circumstances.

“My take on this was that this was one of those successful programs,” Barbara Schneider, who co-edited the Scale-Up in Education series, published by Rowman & Littlefield Publishers of Lanham, Md., said of the IERI. Ms. Schneider is the principal investigator for the Data Research and Development Center at the University of Chicago, a research and technical center set up by the NSF to support the
interagency project.

“Some wonderful research came out of this,” she added, “and some rigorous research that resulted in practical applications for schools.”

**Downsized in Detroit**

Ms. Songer’s eight-year project, now renamed BioKIDS, offers a prime example. Following that first attempt at replicating results on a larger scale, the Michigan researchers decided the way to scale up was to first scale down.

First, they returned to the lab and developed new curriculum units for teaching middle schoolers about biodiversity. As part of the innovative approach, students use Internet resources and handheld animal-tracking devices to inventory the insects, birds, and other signs of wildlife they find in their schoolyards.

The scholars also tried to structure the program in ways they hoped would foster more-complex learning. Students would have to spend more time on the units, for example, and collect data, make scientific claims, and back up their claims with evidence. The shift in focus, though, also meant researchers would have to go beyond traditional standardized tests and devise new assessments to gauge whether students were actually making deeper learning connections.

Then, the researchers cut back the number of teachers and students in their research, choosing instead to focus only on middle schools in Detroit, where passing rates on state science exams lagged far behind the state average. The shift in focus appeared to pay off. The more time students spent on the various curricular units the researchers had created, the more they were able to think like scientists and engage in complex reasoning, the alternative assessments showed.

In a study that tracked 2,000 Detroit 6th graders split into treatment and control groups, the researchers also found that the new units led to bigger gains in science on Michigan’s state exams. On the 8th grade state science exam, the study showed, the improvement had reduced the gap between the passing rate for Detroit middle school students and the statewide average from 30 percent to 20 percent. The teaching units are now officially part of the Detroit district’s science curriculum.

**How Much Control?**

The need to strike a balance between remaining faithful to an original program model and adapting interventions to local circumstances—or at least getting buy-in from the participants—was also a recurring theme across several of the projects.

“I think that’s a very common tension in the whole scale-up process,” said James M. McPartland, who with colleagues at John Hopkins University in Baltimore developed the Talent Development model for improving middle and high schools. “How much do you specify, and how much should be co-created with the locals in the school system? In our own research, we come down on the side of specificity.”

In their chapter in the second volume of the book series, Mr. McPartland and his colleagues present data from surveys of 70,000 students in 59 Talent Development high schools to see how forming smaller schools or academies within larger high schools—one piece of the Talent Development model
—affects the degree to which students feel that teachers care about them or take responsibility for their learning.

What they found was that smaller learning communities seemed to be seven times more effective when paired with other pieces of the model, such as team-teaching or schedules that keep students within their academies for all or most of their classes, than when those elements are not present.

Success for All, another widely used improvement program supported by the Interagency Education Research Initiative, puts an even greater focus on specificity. The reading program seeks local buy-in, however, by only entering schools where 80 percent of faculty members agree, by secret ballot, to adopt the model.

With 16 years’ experience that includes replication in thousands of schools across the country, Success For All may be the granddaddy of scaling-up among the IERI projects. But in his chapter, Robert E. Slavin, the Johns Hopkins University researcher who co-founded the program, describes how the program’s rapid buildup in the 1990s strained researchers’ and the university’s capacity to support it.

“The university was the ideal place to do initial development and research,” Mr. Slavin said in an interview. “Once you get into scale-up, you’re into managing an enterprise that’s very dissimilar from what universities do and there are all sorts of practical problems associated with that.”

One example: Under Johns Hopkins’ pay scale at the time, in the mid- to late-1990s, the most the project could pay a chief executive officer was $40,000 a year, according to Mr. Slavin.

Mr. Slavin was reluctant, though, to turn the program over to any of the for-profit companies that offered to market it for him. His eventual solution was to form a nonprofit foundation, headed by himself and his wife, to continue to steer the program’s growth.

“But most researchers,” Mr. Slavin said, “would not want to do that.”

**Working at Cross-Purposes**

Barbara S. Foorman, an education professor at Florida State University in Tallahassee, describes a rocky experience that her research team had when it collaborated with a for-profit technology firm to study an early-reading-assessment scale developed for Texas public schools.

The plan was to conduct a randomized study in 245 schools to test whether the assessment was more
effective using handheld computers, the Web, or paper-and-pencil forms and to gauge the level of support teachers needed to use it well. The problem was that salespeople for the tech firm continued to peddle a handheld computer to schools that were in the designated control group for the study—a practice that could have jeopardized the multimillion-dollar project had it not been stopped. In the end, though, no test-delivery method proved superior.

“You have this dilemma” with scaling up educational interventions, Ms. Foorman said in an interview. “You have to maintain control, but in doing so you often preclude any real market exposure you have for your product.”

Ms. Foorman’s experiences are described in *Issues in Practice*, the second volume in the *Scale-Up in Education* series. The first volume, which carries the subtitle *Ideas in Principle*, describes the theories underlying the federal interagency scale-up efforts. A third volume, synthesizing ideas and findings from all of the projects, some of which are still going on, is due out a year from now.

*Coverage of education research is supported in part by a grant from the Spencer Foundation.*