Federal policymakers are calling on the education research community to use rigorous methods to evaluate and clearly articulate "What Works?" New legislation requires that federally-funded research in education rely on experimental methods and focus on measuring the main effects of treatments. The intent is to produce research results that yield knowledge about the practices that ought to be scaled up on the basis of scientific research. To support this effort, the administration has funded the What Works Clearinghouse, which will serve as a repository for researchers, policy makers, and educators of research findings from main effects studies.

Although appealing, this seemingly straightforward approach is incomplete as a basis for scaling up educational interventions in general, and those involving sophisticated uses of technology in particular. Scaling up reform necessarily involves adaptation by local actors to local contexts, adaptations that can yield different results in different contexts. Local decision makers need research findings that shed light on expected effects under different circumstances and on the contextual and implementation factors that are likely to influence success.

Scaling of complex educational interventions should be informed by research that sheds light on the more complex question "What Works When and How?" This alternative statement of the problem draws researchers' attention to the need to deal with issues of implementation and context in their designs in a way often neglected by studies focused on main effects. Implementation of technology-supported reforms is typically complex, requiring teachers to orchestrate new technologies in the classroom as they are also introducing new curricular materials and patterns of social interaction into their

teaching. Contexts typically exert a strong influence over implementation in technology-supported reforms; access to technology, reliability of connections to the Internet, and availability of opportunities for teachers to learn about how to integrate technology are all factors that shape implementation. Only studies that gather data on outcomes *and* on context and implementation are in a position to provide local policy makers with crucial information about what works when and how that can inform their decision making.

To illustrate the importance of addressing the implementation and context issues (as well as the importance of having appropriate student outcome measures), we present an example of a technology-supported innovation that has "gone to scale" and some of our research on the program. Since 1995, researchers at the Center for Technology in Learning have been evaluators of the GLOBE Program, an environmental science and education program that has reached over 20,000 teachers in more than 100 countries. As part of our evaluation, we have conducted a number of studies of student learning, including some with comparison groups. We have found in these studies that GLOBE makes a difference in student learning, but only when certain aspects of the program are implemented. To complement our student assessment findings, we have conducted other analyses that demonstrate the influence that specific school site characteristics, technology configuration, and support from local training providers have on implementation of critical aspects of the program. Taken together, these results have provided guidance to GLOBE concerning where to allocate scarce resources to maximize the impact on teaching and learning.

For researchers, the availability of broad and detailed data on context and implementation supports the development of more refined evaluation studies. At the

outset of the evaluation, we focused on simple comparisons of performance of students from GLOBE and non-GLOBE schools on our assessment measures. Over time, we have used implementation data to revise and refine our student learning measures and our techniques for gathering data. We have posed more refined hypotheses about student learning and adopted techniques for statistical modeling of enactment variables to test those hypotheses, without having to adopt a "kitchen-sink" approach to investigating all possible predictors of student learning. On the basis of our experience in GLOBE and in a range of other projects, we suggest that a model of evaluation research that focuses on outcomes, implementation, and contexts simultaneously is critical to scaling-up research on technology-supported innovations.