Title: Multisite Studies and Scaling-Up in Educational Research
Abstract Body

Background / Context:

A scale-up study in education typically expands the sample of students, schools, districts, and/or practices or materials used in the original (smaller) study in ways that build in heterogeneity. The Institute of Education Sciences (IES) website offers the following description of scaling up under **Goal Four**:

**“Goal Four:** If interventions are able to produce positive effects in small efficacy evaluations, they may be ready to be evaluated in a scale-up evaluation. Scale-up evaluations determine whether or not an intervention is effective when it is implemented under conditions that would be typical if the district were to implement it on its own (i.e., without special support from the developer or research team) across a variety of conditions (e.g., different student populations, different types of schools).”

There appears to be agreement in education that scaling up an original study (called a demonstration study in some literatures) for the purpose of evaluating an intervention on a larger scale (e.g., district, state) is not simply a matter of doing the same thing with more resources. Still, there are somewhat different conceptualizations of scaling-up. For example, McDonald, Keesler, Kauffman, and Schneider (2006) stated that

“We view scale-up as inherently about size, numbers, “doing more”—about extending the reach of an exemplary intervention to produce similarly positive effects in different settings and to help a greater number of students. Interventions that are not implemented with larger numbers (of students, teachers) are not “scaled-up”—they are local interventions with promising results.” (p. 16)

Coburn (2003), on the other hand, discussed the importance of scaling up as involving something beyond simply “doing more” and proposed

“conceptualizing scale in four dimensions: “depth, sustainability, spread and shift in reform ownership” (p. 4).

Whatever the conceptualization, it is hard to overstate the need for an effective intervention to be evaluated in a scale-up study or the difficulty of doing so.

Need To Scale Up

Evidence that an educational intervention is effective (e.g., improves student achievement, promotes stronger community-school partnerships) for a few classrooms or schools is not generally sufficient. Many educational researchers have noted that systemic problems in the U.S. educational system require systemic solutions (e.g., Resnick, 2007), which creates a need to
identify interventions that are effective across “…a variety of conditions (e.g., different student populations, different types of schools)” (IES website).

An example of the need to scale up is provided in the 2005 American Educational Research Association panel (Cochran-Smith & Zeichner, 2005). This report amply documented the need to conduct research that evaluates the effectiveness of promising teacher preparation programs on a national scale.

A second and quite different example of the need to scale up is the National Science Founded (NSF)-funded Minnesota Mathematics and Assessment Project (MNMAP) (see Harwell et al., 2009), which examined the impact of the high school mathematics curriculum a student completed on their college mathematics course-taking and achievement using data from a single postsecondary institution. Because the relationship between high school mathematics preparation and student course-taking and achievement may vary across different kinds of postsecondary institutions (e.g., two-year and four-year schools, more or less selective institutions), it is important to learn whether the demonstration study findings generalize (scale-up) to a range of postsecondary institutions.

Difficulty of Scaling Up

Surprisingly little is known about the factors that promote successful scaling up efforts in education, in large part due to the absence of empirically supported theories of scaling up (Sternberg et al., 2006). While there is movement in this area as evidenced by the appearance of scholarly papers (e.g., McDonald et al., 2006), books such as Scale-up in Education: Issues in Practice (edited by B. Schneider and S.K. McDonald, 2006), and the establishment of the National Center on Scaling Up Effective Schools at Vanderbilt University, a core literature that researchers and policymakers interested in scaling-up studies can turn to for guidance is years away. Moreover, developing a literature to guide scaling-up studies represents a significant challenge because of the complexity of this effort. Dewa et al. (2002) captured this reality:

“The proposition of introducing the same study design in different settings and programs is deceptively straightforward. The difficulty is not in the conceptualization but in the implementation.” (p. 173)

Adding to the complexity of identifying factors that promote successful scale-up efforts are the

“…implications such research carries for the work of researchers, teacher educators, and school personnel. They are enormous. Related complexities and their interactions among the development of measures, data collection tasks, and analysis and interpretation demands seem endless.” (Schalock, Schalock, & Ayres, 2006, p. 107) (see also McDermott, 2000 and Sternberg et al., 2006).

A few papers in education have attempted to categorize these challenges and to suggest preliminary guidelines for scale-up studies, but overlap among different categorizations is disappointingly modest. For example, McDonald et al. (2006) argued that context is important and described three main challenges in scaling up research in education: internal and external validity, statistical power and sample size, and identifying methodological tools that produce robust and generalizable findings. On the other hand, Sternberg et al. (2006) also argued that context is important but focused on
characteristics of scaling up that were informed by the work of Brunswik (1956) and included heterogeneity of content and skills standards across states, districts, and schools; heterogeneity of students’ ability levels across and within schools; heterogeneity of teachers’ skills; and embedded accountability of student progress.

Purpose / Objective / Research Question / Focus of Study:

The purpose of the proposed work is to describe how practices, strategies, and models for conducting multisite studies can inform scaling-up a demonstration study in education. The basic idea is that (1) multisite studies often follow a single site (demonstration) study and thus represent a kind of scaling up (2) a literature documenting successful practices, strategies, and models for planning and conducting multisite studies exist (especially in public health) and can be used to inform the planning and conduct of scaling-up studies in education.

Dewa et al (2002) noted that the usual argument for multisite studies is that they allow the same outcomes to be measured with the same instruments using the same timeframe across (possibly) an intervention that may have multiple but comparable versions, and enhance generalizability of findings. Similarly, Raudenbush and Liu (2000) offered several reasons for performing multisite studies. Perhaps the best known multisite study in education is the Tennessee Class Size Study (Finn & Achilles, 1990; Mosteller, 1995). Unfortunately, the surge of interest in multisite studies in education (see, e.g., McDonald et al., 2006; National Research Council, 2004; Raudenbush & Liu, 2000; Zeichner, 2005) has not yet produced a mature literature documenting the successful practices, strategies, and models for planning and conducting these studies that respond to the challenges of scale-up evaluations.

While multisite studies are increasingly popular in education they have been a staple of public health research for decades, and a mature literature is in place. An examination of this literature shows that the challenges in moving from a single site (demonstration) study to a multisite study are similar to those in scaling up a demonstration study in education. Moreover, the public health research literature has identified practices, strategies, and models for multisite studies that respond successfully to many of these challenges including: identifying realistic goals, matching measured outcomes to program specifics, installing and maintaining various lines of communication, support mechanisms for maintaining fidelity of implementation of an intervention, standardizing documents and protocols, managing time and costs, securing cooperative agreements and outsourcing, evaluations of program development, facilitation of community building, providing regular empirical feedback, conducting data analyses that enhance generalizibility arguments, and disseminating results (Constantine & Cagampang, 1998; Lebowitz, Vitiello, & Norquist, 2003; Schalock, Schalock, & Ayres, 2006). These factors offer a blueprint for planning and conducting a scaling up of a demonstration study in education.

Setting:

Two multisite studies in education will be examined using the practices, strategies, and models documented in the public health research literature (Tennessee Class Size Study and MNMAP), as well as those available in the education literature (e.g., Sternberg et al., 2006). The Tennessee Class Size Study
is perhaps the best known educational study conducted in the last 30 years, and there is a wealth of
documentation available for this study. The NSF-funded MNMAP study examined the impact of the high
school mathematics curriculum a student completed on their college mathematics course-taking and
achievement (see Harwell et al., 2009).

Population / Participants / Subjects:

For descriptions of the participants in the Tennessee Class Size Study see Finn and Achilles (1999),
Mosteller (1995) and information on the websites

http://www.heros-inc.org/star.htm#docs and
http://harrisschool.uchicago.edu/about/publications/working-papers/pdf/wp_06_06.pdf. For MNMAP
the population of interest was students enrolled in a four-year or two-year postsecondary institution in
the U.S. All data were archival. MNMAP began with data from a single institution and then expanded
(scaled-up) to a total of 53 four-year and two-year schools and approximately 20,000 students.

Intervention / Program / Practice:

For descriptions of the intervention in the Tennessee Class Size Study see Finn and Achilles (1999). The
intervention in MNMAP is the high school mathematics curriculum a student completed (traditional,
standards-based, or University of Chicago School Mathematics Program, see Harwell et al., 2009).

Significance / Novelty of study:

Multisite studies in education are likely to continue to increase in frequency yet the planning and
conduct of these studies has not been guided by a mature literature documenting successful practices,
strategies, and models. Drawing on literature in public health for planning and conducting multisite
studies in ways that optimize what is learned and its generalizibility represents an important
interdisciplinary effort.

Usefulness / Applicability of Method:

The Tennessee Class Size and MNMAP studies will be used to illustrate the practices, strategies, and
models for planning and conducting multisite studies documented in public health research. This will
provide preliminary guidelines for these studies and, in conjunction with existing scale-up literature,
identify researchable areas.

Conclusions:

Drawing on existing models for conducting multisite studies in public health research can provide
important guidance to educational researchers interested in scaling up a demonstration study. This
interdisciplinary focus may also suggest researchable areas.
Appendix A. References


