Sternberg, R. J., Birney, D., Jarvin, L., Kirlik, A., Stemler, S. & Grigorenko, E. L. (2006). From Molehill to Mountain: The Process of Scaling Up Educational Interventions (First-Hand Experience Upscaling the Theory of Successful Intelligence). In M. A. Constas, & R. J. Sternberg (Eds.) *Translating theory and research into educational practice: Developments in content domains, large-scale reform, and intellectual capacity* (pp. 205–221). Mahwah, NJ: Lawrence Erlbaum Associates.

9

From Molehill to Mountain: The Process of Scaling Up Educational Interventions (Firsthand Experience Upscaling the Theory of Successful Intelligence)

Robert J. Sternberg
Tufts University

Damian Birney
University of Sydney, Australia

Alex Kirlik
University of Illinois at Urbana–Champaign

Steven Stemler
Wesleyan University

Linda Jarvin and Elena L. Grigorenko
Yale University

Embedded into the question of how to translate research into practice are a number of issues, one of which is scale. Once the effectiveness of an intervention has been established at a limited scale in a fairly controlled environment (e.g., a classroom or school), what steps need to be taken to ensure that efforts to translate research into practice at larger scales (e.g., school district, multiple school districts, state, regions), where a high level of control of implementation is not feasible, will be successful? Over the past few years, increasing numbers of researchers have begun to consider how interventions may "scale up."

Some educational interventions successfully scale up; others do not. Little—arguably, almost nothing—is known about the factors that lead to successful upscaling. Our goal in this chapter is to identify a number of these factors through a disciplined and methodologically rigorous approach. In addition, we discuss some lessons learned through our own efforts to upscale an educational intervention based on the theory of successful intelligence (Sternberg 1997, 1999).

This chapter is based on the idea that Brunswik's (1956) theory of probabilistic functionalism, and more specifically, his notion of representativeness, provides a theoretical framework suited to meeting the goals of upscaling and to addressing the problem of identifying the conditions under which educational interventions to improve pre-kindergarten to Grade 12 learning will succeed when applied on a broad scale. To understand scalability, it is necessary to conduct careful investigations guided by the prescriptions of representativeness. These investigations should focus on both the characteristics of educational programs and the descriptions of environmental conditions toward which educational interventions should be targeted.

WHY IS UPSCALING HARD TO DO?

The problem of moving an educational intervention from one location to many locations is beginning to receive considerable attention in the research literature (e.g., Berends, Kirby, Naftel, & McKelvey, 2001; Elmore, 1996; Fullan, 2000; Ramey & Ramey, 1998; Slavin & Madden, 1996). In principle, it might seem that scaling up would be a simple, even trivial, task: One simply takes an intervention that has been used on a small scale and applies it on a large scale. In practice, however, there are many difficulties associated with scaling up. These difficulties might be broadly summarized as falling into two classes: (a) difficulties associated with interventions (i.e., is a particular program suitable for upscaling?) and (b) difficulties associated with the social and human contexts in which interventions are implemented (i.e., what contexts are suitable for upscaling?).

With regard to the first class of difficulties, the first characteristic is obvious—the intervention itself should be of high quality. Elmore (1996) highlights a number of other key features of programs suitable for successful upscaling. For example, the program must create realistic expectations about the amount of time required for teachers and principals to master and deliver the program. Programs often work on a small scale because they are adopted by highly motivated individuals. Yet, these same programs fail to successfully upscale because the initial sample of the program's deliverers was not representative of the larger population. Moreover, the program must provide sufficient scaffolding for teachers navigating and delivering the program for the first time. The program must also be one that teachers can cope with, given their competencies and skills, and that they can implement with a reasonable amount of in-service instruction given the other demands on their time.

With regard to the second class of difficulties, there are also several groups of problematic issues. The first has to do with getting people to adopt the program (i.e., the diffusion of the innovation). The second problem relates to the degree to which the program is successfully implemented. The issue of adoption is relevant to the problem of upscaling because no intervention ever can be upscaled if educators do not buy it. Of relevance here are Rogers's (1995) Diffusion of Innovations and Ely's (1976) Conditions of Change models of educational change. The Diffusion of Innovations framework emphasizes the impact of characteristics of the innovation on the rate of adoption, whereas the Conditions of Change model calls attention to circumstances that predispose an environment toward change.

However, even reasonably good educational interventions might fail the upscaling test if educators do not commit themselves to implementing them (Elmore, 1996). For example, there is some research suggesting that even during peak reform periods, only about 25% of teachers are interested in experimenting with reform efforts (Cuban, 1990; Elmore, 1996). Thus, even a good program might fail to upscale successfully as a result of the resistance of those who are expected to carry out the upscaling.

It is often assumed that a good program or a great idea will sell itself. Perhaps one of the major breakdowns in upscaling happens because good programs have failed to address the critical need to disseminate their findings in a way that communicates effectively with educators. In other words, innovators need to know what educators, who make the critical decisions with regard to bringing innovations to school, consider while these decisions are made. These aspects of the "consumer mentality" of educators in adopting specific interventions have not been well studied. The second issue, implementation, is also relevant to the problem of upscaling. Even when a decision to adopt successful interventions is made, adopting institutions have to consider all kinds of factors (e.g., institutional and individual competence of the program implementers and social and structural conditions of implementations) that influence upscaling (Fullan, 1982; Fullan & Miles, 1992).

In addition, there are a number of themes that have been developed in the organizational psychology literature that link adoption and implementation (Armenakis & Bedeian, 1999; Brink et al., 1995; Goodman & Steckler, 1989; Goodman, Steckler, Hoover, & Schwartz, 1993). These four research themes are (a) content, which pertains to the substance of the change; (b) context, which involves existing forces or conditions in an organization's external and internal environments; (c) process, which addresses actions undertaken during the enactment of intended change; and (d) criterion, which deals with outcomes or markers for tracking the likelihood that necessary behaviors are enacted to achieve the desired changes. A number of ideas from the organizational-change literature have been successfully used in studies of upscaling in schools (e.g., Blumenfeld et al., 2000). Yet, none of these approaches or theories has been adopted in the field of education. Correspondingly, there is a need both to explore developed theories and to propose new approaches.

THEORETICAL FRAMEWORK FOR UPSCALING

One angle from which the scalability issue can be approached is that of scientific generalization (Brunswik, 1956). Reformulated in the context of educational interventions, the issue of scientific generalization can be stated as follows: What characteristics of a given educational program have the potential to transcend the context of discovery (i.e., the immediate context in which the program was developed) and to be validly implemented in contexts unlike the context of discovery? The hypothesis of the existence of such characteristics originates from practices of natural sciences for which there are universal laws applicable to all (or almost all) contexts. However, a growing apprehension exists in the field of educational sciences that many cognitive and social phenomena may never adequately yield to a description or explanation in terms of "universal" characteristics or laws. For example, as a result of extensive research indicating the importance of phonological skills in mastery of reading, many reading-intervention programs based on teaching phonemic skills have been developed. However, not all children respond equally well to the direct teaching, phonics-based approach—some other interventions are needed to address the needs of a diverse group of children with reading difficulties. (For discussion of this issue, see Blachman, 1997.)

Because universal laws in education are lacking, the task of assessing or attempting to ensure the generalizability of educational research products will have to be based instead on the logic of representativeness. An example of such a task would be determining whether a newly proposed educational intervention emerging from a psychological laboratory or a specific limited educational setting will scale up to the realities of the modern educational environment. The term *representativeness* was introduced into discussions of psychological methodology by Brunswik (1956) to indicate the level of similarity between the context in which a scientific investigation is performed and the target class of contexts toward which research products are intended to generalize. As such, the concept of representativeness embraces the notion that outcomes of any learning process are highly dependent on the context in which this learning unfolds (e.g., Sternberg & Wagner, 1994). At the same time, however, the concept of representativeness provides a logical basis for assessing and assuring generalizability, and thus, it plays an important role both in the authors' general theoretical approach and in our specific research plan for contributing knowledge to better enable the upscaling of educational interventions.

Thus, a functionalist program focuses on principles of adaptation to particular environments rather than on context-free mechanisms. Consequently, the corresponding methodology needs to include characteristics of educational interventions, students to whom these interventions are administered, and the environmental conditions under which these programs are run (Hammond & Stewart, 2001). Because the environments in which schooling and learning take place vary and because schooling is "applied" to all kinds of children at different levels of abilities and of different backgrounds, the generalizations should be based on statistical sampling of the environments as well as on statistical sampling of participants in the process of schooling. The latter is common today: Educational studies routinely employ large numbers of participants to ensure reliable generalization of results beyond the particular participants studied to the wider population to which the investigator intends to generalize his or her results. What is still rather rare is statistical sampling of the environmental conditions toward which generalization is intended. Just as in participant sampling, a crucial factor determining the success of a representative design is accurately specifying and describing the properties of the environmental (in the context of this chapter, educational) settings toward which the research results are intended to apply (or again, to scale up). As with participant sampling, the specification of the target environments toward which generalization is desired should be performed prior to the performing a study.

In our own work, we have used many of the principles described previously, starting with small-scale studies and then upscaling them in what we believe are appropriate ways.

LESSONS FROM A SCALE-UP OF THE THEORY OF SUCCESSFUL INTELLIGENCE: WHAT IS SUCCESSFUL INTELLIGENCE?

The Definition of Successful Intelligence

Intelligence Is Defined in Terms of the Ability to Achieve Success in Life in Terms of One's Personal Standards, Within One's Sociocultural Context. The field of intelligence has at times tended to put the cart before the horse, defining the construct conceptually on the basis of how it is operationalized rather than vice versa. This practice has resulted in tests that stress the academic aspect of intelligence, as one might expect given the origins of modern intelligence testing in the work of Binet and Simon (1905/1916) in designing an instrument that would distinguish children who would succeed from those who would fail in school. However, the construct of intelligence needs to serve a broader purpose, accounting for the bases of success in all areas of one's life.

The use of societal criteria of success (e.g., school grades, personal income) can obscure the fact that these operationalizations often do not capture people's personal notions of success. Some people choose to concentrate on extracurricular activities such as athletics or music and pay less attention to grades in school; others may choose occupations that are meaningful to them but that never will yield the income they could gain doing work that is less personally meaningful. Although scientific analysis of some kinds requires nomothetic operationalizations, the definition of success for an individual is idiographic. In the theory of successful intelligence, however, the conceptualization of intelligence is always within a sociocultural context. Although the processes of intelligence may be common across such contexts, what constitutes success is not. Being a successful member of the clergy of a particular religion may be highly rewarded in one society and viewed as a worthless pursuit in another culture.

One's Ability to Achieve Success Depends on One's Ability to Capitalize on One's Strengths and Correct or Compensate for One's Weaknesses. Theories of intelligence typically specify some relatively fixed set of abilities, whether one general factor and a number of specific factors (Spearman, 1904), seven multiple factors (Thurstone, 1938), eight multiple intelligences (Gardner, 1983, 1999), or 150 separate intellectual abilities (Guilford, 1982). Such a nomothetic specification is useful in establishing a common set of skills to be tested. However, people achieve success, even within a given occupation, in many different ways. For example, successful teachers and researchers achieve success through many different combinations of skills rather than through any single formula that works for all of them.

A Balance of Abilities Is Achieved to Adapt to, Shape, and Select Environments. Definitions of intelligence have traditionally emphasized the role of adaptation to the environment ("Intelligence and Its Measurement," 1921; Sternberg & Detterman, 1986). However, intelligence involves not only modifying oneself to suit the environment (adaptation) but also modifying the environment to suit oneself (shaping) and sometimes finding a new environment that is a better match to one's skills, values, or desires (selection).

Not all people have equal opportunities to adapt to, shape, and select environments. In general, people of higher socioeconomic standing tend to have more opportunities than do people of lower socioeconomic standing.

The economy or political situation of the society also can be factors. Other variables that may affect such opportunities are education and especially literacy, political party, race, religion, and so forth. For example, someone with a college education typically has many more possible career options than does someone who has dropped out of high school to support a family. Thus, how and how well an individual adapts to, shapes, and selects environments must always be viewed in terms of the opportunities the individual has.

Success Is Attained Through a Balance of Analytical, Creative, and Practical Abilities. Analytical abilities are the abilities primarily measured by traditional ability tests. However, success in life requires one not only to analyze one's own ideas and the ideas of others but also to generate ideas and to persuade other people of their value. This necessity occurs in the work world as when subordinates try to convince superiors of the value of their plan, in the world of personal relationships as when children attempt to convince their parents to do what the children want or when spouses try to convince their partners to do things their preferred way, and in the world of school as when students write an essay arguing for a particular point of view.

According to the proposed theory of human intelligence and its development (Sternberg, 1980, 1984, 1985, 1990, 1997, 1999), a common set of processes underlies all aspects of intelligence. These processes are hypothesized to be universal. For example, although the solutions to problems that are considered intelligent in one culture may be different from the solutions considered to be intelligent in another culture, the need to define problems and translate strategies to solve these problems exists in any culture.

Metacomponents, or executive processes, plan what to do, monitor things as they are being done, and evaluate things after they are done. Examples of metacomponents are recognizing the existence of a problem, defining the nature of the problem, deciding on a strategy for solving the problem, monitoring the solution of the problem, and evaluating the solution after the problem is solved.

Performance components execute the instructions of the metacomponents. For example, inference is used to decide how two stimuli are related, and application is used to apply what one has inferred (Sternberg, 1977). Other examples of performance components are comparison of stimuli, justification of a given response as adequate although not ideal, and actually making the response.

Knowledge-acquisition components are used to learn how to solve problems or simply to acquire declarative knowledge in the first place (Sternberg, 1985). Selective encoding is used to decide what information is relevant in the context of one's learning. Selective comparison is used to bring old information to bear on new problems, and selective combination is used to put together the selectively encoded and compared information into a single and sometimes insightful solution to a problem.

Although the same processes are used for all three aspects of intelligence universally, these processes are applied to different kinds of tasks and situations depending on whether a given problem requires analytical, creative, or practical thinking or a combination of these kinds of thinking. In particular, analytical thinking is invoked when components are applied to fairly familiar kinds of problems abstracted from everyday life. Creative thinking is applied when the components are applied to relatively novel kinds of tasks or situations. Practical thinking is used when the components are applied to experience to adapt to, shape, and select environments.

More about the theory can be found in Sternberg (1985, 1997). Because the theory of successful intelligence comprises three subtheories—a componential subtheory dealing with the components of intelligence; an experiential subtheory dealing with the importance of coping with relative novelty and of automatization of information processing; and a contextual subtheory dealing with processes of adaptation, shaping, and selection—the theory has sometimes been referred to as triarchic.

EDUCATIONAL INTERVENTIONS BASED ON SUCCESSFUL INTELLIGENCE

In early studies (Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999), the question of whether conventional education in school systematically discriminates against children with creative and practical strengths has been explored. Motivating this work was the belief that the systems in most schools strongly tend to favor children with strengths in memory and analytical abilities. However, schools can be unbalanced in other directions as well. One school R. Sternberg and E. L. Grigorenko visited in Russia in 2000 placed a heavy emphasis on the development of creative abilities—much more so than on the development of analytical and practical abilities. While on this trip, they were told of yet another school—catering to the children of Russian businessman—that strongly emphasized practical abilities and where children who were not practically oriented were told that eventually, they would be working for their classmates who were so oriented.

We used the Sternberg (1993) Triarchic Abilities Test, which measures analytical, creative, and practical abilities, in some of our instructional work. The test was administered to 326 children around the United States and in some other countries who were identified by their schools as gifted by any standard whatsoever. Children were selected for a summer program in (college-level) psychology if they fell into one of five ability groupings: high analytical, high creative, high practical, high balanced (high in all three abilities), or low balanced (low in all three abilities). Students who came to Yale were then divided into four instructional groups. Students in all four groups used the same introductory-psychology textbook (a preliminary version of Sternberg, 1995) and listened to the same psychology lectures. What differed among them was the type of afternoon discussion section to which they were assigned. They were assigned to an instructional condition that emphasized either memory, analytical, creative, or practical instruction. For example, in the memory condition, they might be asked to describe the main tenets of a major theory of depression. In the analytical condition, they might be asked to compare and contrast two theories of depression. In the creative condition, they might be asked to formulate their own theory of depression. In the practical condition, they might be asked how they could use what they had learned about depression to help a friend who was depressed.

Students in all four instructional conditions (Sternberg et al., 1999) were evaluated on their performance on homework, a midterm exam, a final exam, and an independent project. Each type of work was evaluated for memory, analytical, creative, and practical quality. Thus, all students were evaluated in exactly the same way. The results suggested the utility of the theory of successful intelligence. This utility showed itself in several ways.

First, students in the high-creative and high-practical groups were much more diverse in terms of racial, ethnic, socioeconomic, and educational backgrounds than were students in the high-analytical group, suggesting that correlations of measured intelligence with status variables such as these may be reduced by using a broader conception of intelligence. Thus, the kinds of students identified as strong differed in terms of populations from which they were drawn in comparison with students identified as strong solely by analytical measures. More important, just by expanding the range of abilities measured, Sternberg et al. (1999) discovered intellectual strengths that might not have been apparent through a conventional test.

Second, Sternberg et al. (1999) found that all three ability tests—analytical, creative, and practical—significantly predicted course performance. When multiple-regression analysis was used, at least two of these ability measures contributed significantly to the prediction of each of the measures of achievement. Perhaps as a reflection of the difficulty of deemphasizing the analytical way of teaching, one of the significant predictors was always the analytical score. (However, in a replication of Sternberg et al.'s [1999] study with low-income African American students from New York, Coates [personal communication, 1997] of the City University of New York found a different pattern of results. Coates's data indicated that the practical tests were better predictors of course performance than were the analytical measures, suggesting that what ability test predicts what criterion depends on population as well as mode of teaching.)

Third and most important, there was an aptitude–treatment interaction whereby students who were placed in instructional conditions that better matched their pattern of abilities outperformed students who were mismatched. In other words, when students are taught in a way that fits how they think, they do better in school. Children with creative and practical abilities, who are almost never taught or assessed in a way that matches their pattern of abilities, may be at a disadvantage in course after course, year after year.

Follow-up studies (Sternberg, Torff, & Grigorenko, 1998a, 1998b) have examined learning of social studies and science by third graders and eighth graders. The 225 third graders were students in a very low-income neighborhood in Raleigh, North Carolina. The 142 eighth graders were largely middle- to upper middle-class students in Baltimore, Maryland, and Fresno, California. Students were assigned to one of three instructional conditions. In the first condition, they were taught the course they would have learned had there been no intervention. The emphasis in the course was on memory. In a second condition, students were taught in a way that emphasized critical (analytical) thinking. In the third condition, they were taught in a way that emphasized analytical, creative, and practical thinking. All students' performance was assessed for memory learning (through multiple-choice assessments) as well as for analytical, creative, and practical learning (through performance assessments).

As expected, students in the successful-intelligence (analytical, creative, practical) condition outperformed the other students in terms of the performance assessments. One could argue that this result merely reflected the way they were taught. Nevertheless, the result suggested that teaching for these kinds of thinking succeeded. More important, however, was the result that children in the successful intelligence condition outperformed the other children even on the multiple-choice memory tests. In other words, to the extent that one's goal is just to maximize children's memory for information, teaching for successful intelligence is still superior. It enables children to capitalize on their strengths and to correct or compensate for their weaknesses, and it allows children to encode material in a variety of interesting ways (Sternberg, Torff, & Grigorenko, 1998a, 1998b).

Thus, the results of three sets of studies suggest that the theory of successful intelligence is valid as a whole. Moreover, the results suggest that the theory can make a difference not only in laboratory tests but in school classrooms and even the everyday life of adults as well.

More recently, Grigorenko, Jarvin, and Sternberg (2002) reported on three studies that extended the work on applying the theory of successful intelligence in the classroom with the goal of improving reading performance. As in the earlier studies, Grigorenko et al. attempted to help teachers do better what they were already doing (e.g., teaching reading) rather than giving them a new curriculum that they most likely would reject for lack of time. Hence, in these studies, the goal was to supplement standard reading instruction—which included both phonic and whole-language elements—with a specifically creative, analytical, practical (CAP) intervention.

In brief, Grigorenko et al. (2002) worked with the reading curricula at the middle and high school levels. In this study of 871 middle school students and 432 high school students, Grigorenko et al. taught reading either triarchically or through the regular curriculum. At the middle school level, reading was taught explicitly. At the high school level, reading was infused into instruction in mathematics, physical sciences, social sciences, English, history, foreign languages, and the arts. Teachers in the control condition were shown how to apply mnemonic strategies.

To illustrate, in one of the studies Grigorenko et al. (2002) conducted with fifth graders, vocabulary and comprehension were assessed on multiple occasions throughout the academic year. Students in the successful intelligence condition improved over time in memory–analytical, practical, and creative tasks. Students in the control condition displayed a very different performance profile over time: They did not show improvement on memory–analytical tasks, improved on practical tasks, and declined on creative tasks. Hence, students taught using triarchic (successful intelligence) methods profited more over time from instruction than students not taught this way. Grigorenko et al.'s Studies 2 and 3 explored the impact of triarchic teaching on reading in broader academic contexts with upper middle school and high school students. Once again, a distinct triarchic advantage was demonstrated in these samples. In all settings, students who were taught triarchically substantially outperformed students who were taught in standard ways (Grigorenko et al., 2002).

In our Interagency Education Research Initiative-funded project, which was aimed at upscaling the earlier work, we created Grade 4 curriculum materials in three subject areas—language arts (5 units), mathematics (5 units), and science (4 units). The materials were designed for presentation in one of three instructional modes: triarchic (CAP), critical thinking (CT), and memory (M). Across the 14 curricular units, multiple assessments from 7,702 students were collected in the first 2 years of the program. Here, we present the preliminary findings based on a subset of the data. Again, the triarchic advantage is evident. For example, in the "Wonders of Nature" language arts unit, analyses of data showed that significantly greater gains over time were achieved by CAP students compared with both CT and M students. Similar results were found in an analysis of the available data in the "Equivalent Fractions" math unit: CAP students had significantly higher levels of gain than did students in the CT condition and tended to outperform the M students, although the effect vis-à-vis memory has not reached $p < .05$ significance for the subset of data analyzed.

Taken as a whole, the data collected at the elementary, middle, and high school levels are supportive of the utility of the triarchic model of instruction and justify further explorations into the scalability of the triarchic pedagogical approach.

FACTORS ASSOCIATED WITH SUCCESSFUL UPSCALING: LESSONS FROM THE FIELD

In our own research, we have identified some of the empirical factors that can make successful upscaling a challenge. Here we list some of them briefly:

1. Heterogeneity of content and skills standards across states, districts, and schools.
2. Heterogeneity of students' ability levels across and within schools.
3. Heterogeneity of districts: (a) political environment, (b) commitment, and (c) teachers/administrators' levels of experience.
4. Large-scale training of teachers, especially with regard to (a) scheduling, (b) availability, and (c) comparability of training within the groups of trained teachers with regard to (a) distances, (b) budget, (c) quality control, and (d) confidentiality.
5. Lack of control of intervention: (a) experimental (randomization), (b) administrative, (c) accountability, (d) implementation, (e) delivery of materials, and (f) communication with teachers.
6. Variability in technological resources across schools.
7. Equating achievement indicators across schools.
8. Quantification of the outcomes of the programs.
9. Developing assessments of program outcomes and understanding variables that influence assessment task difficulty: (a) ability to generate new tasks and (b) ability to generate alternate forms of the same tasks.
10. Complexity of data analysis: (a) comparability of information and data available for comparable analyses and (b) sources of variability that are controlled and uncontrolled across samples.

Although these factors have been elicited empirically from our own experience in the field, there is a body of general literature on innovation supporting their importance (e.g., Rogers, 1995). Specifically, Veir (1990) used stepwise regression analysis to identify the key context variables for implementing staff-development programs in rural schools. Eight predictor variables were found to be important in explaining whether the program would be successfully implemented in schools. The variables were training time, socioeconomic profile of the student body, administrative participation, proximity to an institution of higher education, provision of incentives, number of high school level teachers, number of administrators in the district, and the presence of a trainer from a higher educational institution.

As we indicated at the onset of the chapter, we find the concept of representativeness especially valuable because it embraces all the empirical observations we made while trying to upscale the successful intelligence intervention. The absolute and relative contributions of all the factors just listed have yet to be quantified, but when considered simultaneously, they provide enough dimensions to compare contexts in which innovative interventions are to be applied on a large scale so that the success of the application can be predicted.

CONCLUSION

Earlier, we noted that educational contexts are characterized by high levels of heterogeneity along many dimensions, creating a challenge to successful upscaling. This, coupled with a theory of successful intelligence acknowledging that learning outcomes are highly context dependent, leads us directly to the view that there is there is simply no magic bullet for the upscaling problem other than to sample and represent this heterogeneity in empirical research. We hope that the diversity and progression of the studies described in this chapter amply illustrate this point.

As evidenced by our framing of upscaling in terms of the more general issue of statistical generalization, it should not come as a surprise that we are not the first to note the need to sample diverse contexts in educational science. More than 60 years ago, in his text on statistical methods for educational research, Lindquist (1940) suggested that interventions be evaluated using a random selection of schools. Nor is the upscaling problem specific to education: Educational scientists must draw on lessons learned by those performing practically relevant research in neighboring psychological disciplines. Successful upscaling is a challenge any time theoretically motivated interventions are applied in specific contexts that were not the primary empirical foundation for theory. Consider, for example, what the human factors pioneer Chapanis (1988) had to say on this point:

> There are two ways one can go about doing studies that will extrapolate to a wide variety of situations. The first is by deliberately building heterogeneity into studies, a tactic recommended by Brunswik (1956) over 30 years ago; the second is by replication. (pp. 253–267).

As scientists striving for robustness and generality in our educational research, we are clearly not alone in embracing Brunswik's (1956) method of representative design as a basis for the design of effective psychological interventions. We believe achieving the goal of generality or robustness is best achieved by iteratively creating, testing, modifying, and retesting theory in heterogeneous contexts.

Representative design solves the problem of generalization by requiring that the integrity and heterogeneity of contextual variables in the target educational environment be preserved in one's fundamental research. To the extent that our discussion and examples have persuaded others to join us in this venture, we note that representative design places one additional requirement on research design and communication. This requirement concerns the need to treat the educational ecology as an object of scientific study on a par with internal psychological activities or processes.

Embracing this aspect of representative design requires an appreciation of Brunswik's (1956) deep understanding of the inherently social nature of scientific conduct and progress. Brunswik noted that truly cumulative knowledge about the contextual influences on cognition and behavior will be achievable only if those conducting and reporting research provide descriptions or theories of their research contexts as precise and detailed as the descriptions or theories psychologists have provided for phenomena such as learning, intelligence, and the like. By requiring this type of "equal treatment" for theories of internal and contextual aspects of cognition and behavior, empirical findings can bear not only on the truth or utility of related phenomena such as learning and intelligence but also on the truth or utility of theory related to the contextual influences on these activities.

By making contextual descriptions and theories explicit in this way, the educational science community as a whole can participate in what Brunswik (1956) called the large, concerted group project that is necessary to realize the promised fruit of representative design. For educational science, this fruit would be a detailed understanding of what contextual factors matter in upscaling and applying interventions, what factors tend not to matter, and what factors matter in what ways. Naturally, these factors may be intervention specific, with the result that what is learned through research is which types of interventions work in which contexts.

When the contextual theories and descriptions underlying educational research are explicitly stated and communicated and findings shared, questions about the robustness and generality of those findings become empirically decidable and thereby scientific. One researcher, for example, may not find a particular effect in an educational environment in which a theory's contextual components suggest she or he should find one. This evidence can then be used to revise the contextual theory, growing one's collective understanding of the contextual variables that matter. We encourage others to join us in this effort, as we believe it to be the most promising path toward improving the scalability of educational interventions as we work to translate research and theory into practice across a diverse range of settings, and upscale and disseminate the most successful and effective research and theory.

ACKNOWLEDGMENTS

Preparation of this article was supported by Grant REC–9979843 from the National Science Foundation. Grantees undertaking such projects are encouraged to express freely their professional judgment. This article, therefore, does not necessarily represent the positions or the policies of the U.S. government, and no official endorsement should be inferred. We thank Cynthia Matthew for review of the relevant organizational psychology literature.

REFERENCES

Armenakis, A. A., & Bedeian, A. G. (1999). Organizational change: A review of the theory and research in the 1990's. *Journal of Management, 25,* 293–315.

Berends, M., Kirby, S., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: Rand Education.

Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins. (Original work published 1905)

Blachman, B. (Ed.). (1997). *Foundations of reading acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist, 35*(3), 149–164.

Brink, S. G., Basen-Engquist, K. M., O'Hara-Tompkins, N. M., Paviel, G. S., Gottlieb, N. H., & Koyato, C. Y. (1995). Diffusion of an effective tobacco prevention program. Part I: Evaluation of the dissemination phase. *Health Education Research, 10,* 283–295.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Chapanis, A. (1988). Some generalizations about generalization. *Human Factors, 30,* 253–267.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3–13.

Elmore, D. R. (1996). Getting to scale with good educational practices. *Harvard Educational Review, 66,* 1–26.

Ely, D. P. (1976). Creating the conditions for change. In S. Fabisoff & G. Bonn (Eds.), *Changing times, changing libraries* (pp. 150–162). Champaign: University of Illinois Graduate School Library.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers College Press.

Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change, 1,* 1–23.

Fullan, M. G., & Miles, M. B. (1992). Getting reform right: What works and what doesn't. *Phi Delta Kappan, 73,* 744–752.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1999). Are there additional intelligences? The case for naturalist, spiritual, and existential intelligences. In J. Kane (Ed.), *Education, information, and transformation* (pp. 111–131). Upper Saddle River, NJ: Prentice Hall.

Goodman, R. M., & Steckler, A. (1989). A model for the institutionalization of health promotion programs. *Family and Community Health, 11,* 63–78.

Goodman, R. M., Steckler, A., Hoover, S., & Schwartz, R. (1993). A critique of contemporary community health promotion approaches: Based on a qualitative review of six programs in Maine. *American Journal of Health Promotion, 7,* 208–220.

Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2002). School-based tests of the triarchic theory of intelligence: Three settings, three samples, three syllabi. *Contemporary Educational Psychology, 27,* 167–208.

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89,* 48–59.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications.* Oxford, England: Oxford University Press.

Intelligence and Its Measurement: A symposium. (1921). *Journal of Educational Psychology, 12,* 123–147, 195–216, 271–275.

Lindquist, E. F. (1940). *Statistical analysis in educational research.* Boston: Houghton Mifflin.

Ramey, C., & Ramey, S. (1998). Early intervention and early experience. *American Psychologist, 53,* 109–120.

Rogers, E. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Slavin, R. E., & Madden, N. A. (1996, November). *Scaling up: Lessons learned in the dissemination of Success For All.* Retrieved April 13, 2002, from http://www.csos.jhu.edu/crespar/Reports/report06entire.html

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review, 84,* 353–378.

Sternberg, R. J. (1980). Sketch of a componential subtheory of human intelligence. *Behavioral and Brain Sciences, 3,* 573–584.

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. *Behavioral and Brain Sciences, 7,* 269–287.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1993). Sternberg Triarchic Abilities Test. Unpublished manuscript.

Sternberg, R. J. (1995). *In search of the human mind.* Orlando, FL: Harcourt Brace College Publishers.

Sternberg, R. J. (1997). *Successful intelligence.* New York: Plume.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology, 3,* 292–316.

Sternberg, R. J., & Clinkenbeard, P. R. (1995). The triarchic model applied to identifying, teaching, and assessing gifted children. *Roeper Review, 17,* 255–260.

Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?* Norwood, NJ: Ablex.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P. R., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly, 40*(3), 129–137.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment, 15,* 1–11.

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998a). Teaching for successful intelligence raises school achievement. *Phi Delta Kappan, 79,* 667–669.

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998b). Teaching triarchically improves school achievement. *Journal of Educational Psychology, 90,* 374–384.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1994). *Mind in context.* New York: Cambridge University Press.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Veir, C. (1990). *Context variables for successful staff development in rural schools.* (ERIC Document Reproduction Service No. ED331184)