

## Impediments to Scaling Up Effective Comprehensive School Reform Models

---

by Siegfried E. Engelmann and Kurt E. Engelmann

Chapter 4 (pages 107–133) in RAND  
*Expanding the Reach of Education Reforms:  
Perspectives from Leaders in the  
Scale-Up of Educational Interventions*

Scaling up involves growth, and any growing organization encounters various impediments as a function of its growth. More people must be recruited, trained, and deployed. Procedures for internal and external communication must be modified. The expanded activities must be funded, which means that the organization must secure additional financial resources. Changes in basic political interactions, public relations, and the organization's goals or structures may also function as impediments to growth.

This chapter does not address these issues—they are growing pains of any organization. Rather, the chapter focuses on impediments unique to organizations that provide services to schools and school districts that have a serious need for models of how to reverse school failure. This chapter focuses on three central issues of scaling up comprehensive reform “designs” or models:

- the relationship between the characteristics of a design and the ease of scaling up the design
- patterns of scaling up and their relative efficiency and economy, with an emphasis on the model-cluster pattern of scaling up
- obstacles inherent in many school districts, especially large urban districts, that hinder the dissemination and implementation of effective models of school reform.

Our identification of these issues grows out of our experience in implementing the Direct Instruction (DI) model of comprehensive school reform.

DI is based on the assumption that controlling the characteristics of instruction and relevant variables in the school setting can greatly accelerate academic achievement. This acceleration is needed most in failed schools, in which high numbers of students perform in the first quartile. DI assumes that even very low-performing children are capable of learning if the instruction is appropriate. The model further assumes that unsatisfactory performance is a direct effect of inadequate teaching.

Compared to traditional programs, the DI model is designed to teach more in the same amount of time. All instructional materials for reading, language, math, spelling, and cultural literacy are specified as required components of the model. These programs, which are commercially available through SRA/McGraw-Hill, are designed to accelerate the performance of children through efficiency, economy, and clarity of presentation. Exercises are scripted so that teachers use effective wording and can present tasks at a relatively high rate. Each program is designed to introduce only a small amount of new material each lesson (about 10 percent of the lesson) with the rest of the lesson continuing the teaching introduced in the previous two lessons and reviewing or applying all the material that children are assumed to have mastered. Each lesson has six to ten ongoing topics that are combined to create increasingly complex applications. The criterion for presenting the material successfully is that, at the end of each lesson, all children are expected to have mastered everything in the current lesson.

The model also requires practices that increase the likelihood that the curricular materials will be used effectively to accelerate student performance by requiring all teachers to participate in the program, providing extensive preservice and in-service training on specific skills and content, and specifying schedules for subjects that provide sufficient daily exposure to the topics being studied. Further, the model specifies grouping children homogeneously, collecting data on student performance, implementing schoolwide management and reinforcement procedures, and using a problem-identification and problem-solution approach.

The basic DI model has undergone extensive evaluation, and, in various meta-analyses, it has been found to be one of the few models for which there is ample evidence of effectiveness—for all populations and socioeconomic levels (Adams and Engelmann, 1996; Borman et al., 2002). In the Follow Through evaluation, the DI model was most successful in all subjects tested (reading, language, math, and spelling), in both basic and cognitive skills (Stebbins et al., 1977). DI also produced the most positive affective outcomes, with DI children having more positive self-images than the children in any other model (Engelmann et al., 1988; Stebbins et al., 1977). The effectiveness of DI has been verified through a variety of other studies addressing a broad range of skills.<sup>1</sup>

### **Assumptions About the Context and Nature of Scaling Up Comprehensive School Reform Models**

The three central issues of scaling up rest on assumptions about models and designs of comprehensive school reform. First, scaling up assumes a model that consists of a stable set of specified or implied practices. As the model is used in an increasing number of schools, the core aspects of the model do not change. Peripheral aspects of the

---

<sup>1</sup> See, for example, Gersten and Keating (1987); Gersten, Keating, and Becker (1988); Carlson et al. (2002). Adams and Engelmann (1996) contains an extensive bibliography of studies on DI.

design may be modified or added to in response to feedback, but the core of the design remains stable.

Second, the logic of schoolwide reform is that schools have been identified as failed schools solely on the basis of student performance. The only possible evidence that a reform model is successful, therefore, is improved student performance. The teachers and administration of the school have not been successful in teaching academic skills—reading, language, math, science, writing. The model may approach this problem either directly or indirectly. Regardless of how circuitous or direct the model is, the test of the model’s effectiveness must be referenced to student success. Observations in the classrooms in which the model has been implemented must disclose that the teachers are teaching age-appropriate academic skills and that the students are learning these skills. Certainly, improved standardized test scores imply that the interactions have changed, but nothing more indirect than standardized test scores would be relevant to whether the model is a model of relevant reform or simply of doing things a different way.

Third, not addressing the classroom as the locus of reform implies that at-risk students are incapable of improving unless something that has nothing to do with the classroom changes. This is a serious philosophical indictment of the students’ presumed capabilities. Enlisting the parents, the business community, and other possible influences is certainly an option, so long as these activities are not simply ends in themselves but lead to observable improvements in student performance in the classroom. That teachers now collaborate and discuss instructional matters is interesting but irrelevant. If classroom interactions do not change in ways that make it obvious that students are learning more in less time, the model is not effectively addressing the problems that spawned the need for models.

Perhaps the most fundamental question is whether models that fail to produce evidence of consistent improvement of student performance *should* be scaled up. It would seem contradictory to scale up any model that does not meet the fundamental requirement of achieving schoolwide *reform* and instead simply produces changes in school-related activities or personnel.

## **Problems in Scaling Up as a Function of the Demands of the Model**

The context and problems of scaling up a reform model are greatly influenced by the design of the model. Different designs have different implications for patterns of scaling up, degrees of resistance from the district, and types of training practices needed for successful implementation.

Each demand that requires the school to do something it is not already doing and that some staff members would not like to do serves as a basis for the school to reject the model. If a model has only one or two such requirements, there would be only one or two possible reasons for the school to reject the model, which means that the “palatability” of the model would be relatively high. In contrast, if a model has 30 such requirements, a school would be far more likely to reject the model.

We can illustrate this relationship with two extreme models for reforming at-risk schools. Let us say that one, called the minimal-requirements model, is based on the proposition that teachers know how to improve student performance and that empowering the teacher with more resources will enable them to improve student performance. The model gives each teacher and the principal in the school a fairly large sum of money. The teachers are required to answer a series of simple questions, attend three motivational sessions, “spend the money wisely,” and fill out an end-of-year summary of how the money was spent and its effects on the school.

The other model, called the extensive-requirements model, addresses all details of how students, teachers, and principals in the failed school interact—what they do and how and when they do it. No detail of the school’s delivery system for academic learning remains unchanged. Furthermore, under the extensive-requirements model, student progress is monitored extensively, and serious attempts are made to ensure that all school practices are congruent with the specifications of the model.

The minimal- and extensive-requirements models differ considerably on five critical variables that influence the potential for scaling

up: training, time required for implementation, effective patterns of scaling up, compatibility with classroom practices, and compatibility with district practices.

### **The Minimal-Requirements Model**

The minimal-requirements model has a much greater potential for scaling up for the following reasons.

**Training.** A very large segment of the available population could serve as “trainers” or facilitators to implement the model. Trainers would only have to demonstrate to school staff how to fill out an end-of-year summary form, how to contact the sponsor, and similar details.

**Time Required for Implementation.** The criteria for full implementation of the model at a school could be met within one year. If all the teachers accept the money, attend the sessions, and fill out the end-of-year report, the model is fully implemented by the end of the year. (The criterion of using the money wisely is general enough that the teachers could meet it by doing almost anything short of activities that are clearly illegal or immoral.)

**Effective Patterns of Scaling Up.** Various patterns of diffusing a minimal-requirements model would be possible because the performance of each implemented school would have little relationship to neighboring schools. Implementing the model in a particular pattern would not be important because various patterns would be equally effective. The geographic proximity of implemented schools would make very little difference. Training could occur remotely through phone, video, or the Internet. The model could be implemented in single schools in a large district as easily as in clusters.

**Model Compatibility with Classroom Practices.** A minimal-requirements model would be compatible with any instructional or management procedure the teachers currently use. Because the model does not demand any type of uniformity or teacher-student achievement, nothing would require teachers to change the way they teach, group students for instruction, or coordinate the practices and curricula from one classroom to the next. The teachers could either do things the way they always have or do them in a different way.

**Compatibility of the Model with District Practices.** A minimal-requirements model would be compatible with a full range of district standards, guidelines, and policies. The model would not affect any of the interactions between schools and the central administration. Whatever instructional requirements had been in place before the introduction of the model could remain in place. The model could readily accommodate any requirement of the district and the presiding teachers' union.

In summary, training time and requirements for the minimal-requirements model would be slight, both for trainers and teachers.

The model could be fully implemented in a single school year and effectively diffused through various patterns, and the model has a very slight potential for creating conflict with teachers, principals, or districts.

### **The Extensive-Requirements Model**

The extensive-requirements model is the antipode of the minimal-requirements model with respect to all issues relevant to scaling up.

**Training.** A very small segment of the available population could serve as “trainers” or facilitators for the model. A trainer would have to be familiar with the various provisions of the model. Because an extensive-requirements model covers a vast range of details and requires performance and monitoring with respect to each detail, the population of potential trainers is very small. Even teachers who have taught the required programs for years could not become trainers until they learned how to train and monitor others, which involves identifying various implementation problems and providing effective solutions.

**Time Required for Implementation.** Extensive-requirements models could not be fully implemented in less than two years and might require as much as five or six years, because teaching teachers to do just about everything in a new way requires extensive professional development. The implementation would occur in stages, so the teachers and principals would be required to implement only some aspects of the model this year and some the next, until full implementation is achieved. Such staging is both efficient and practi-

cal. The teachers will learn only so much during a school year. If they have to learn three new basic curricular sequences (one for reading, one for language, and one for math) and all the related behaviors, many teachers would become overwhelmed. Also, it would be difficult to find the in-service time necessary to bring them up to performance standards. (The greater the amount of school time devoted to in-service, the less time the teachers are working in the classroom, so the less the model is actually being implemented.)

**Effective Patterns of Scaling Up.** Savings in training time and gains in effectiveness are implied by particular diffusion patterns. For the extensive-requirements model, any possible savings are attractive because of the time and cost of implementation. Implementing the model in a geographical cluster of schools could permit efficient staging of the implementation. First, the model would be implemented in a few of the schools. Then one or more of these schools would serve as a training and dissemination center. Selected teachers from the centers would serve as training assistants or coaches in neighboring schools where the model is being implemented. Particularly great savings would be realized during the first two years of implementation at the neighboring schools.

In contrast to the minimal-requirements model, the sponsors of extensive-requirements models would prefer not to implement the model in isolation. When the model is implemented in single, isolated schools, the potential for efficient expansion is limited. Implementation of the model in the tenth geographically remote school would require as much time and effort as in the first. For rural schools and small districts, isolated implementations may be the only option. The greatest savings are possible in large urban districts.

**Model Compatibility with Classroom Practices.** An extensive-requirements model would probably be incompatible with a wide range of instructional and management procedures that teachers currently use. The model would specify exactly what teachers teach, exactly how they teach it, the daily schedule for teaching, the management practices they follow, the way they group students for instruction, and the records they keep about student performance.



This endeavor generates a litany of possible conflicts. Some teachers may prefer using their current instructional materials; some may prefer a different schedule for teaching various subjects; some may prefer a different method of interacting with the students, a different grouping format, or a different method of assessing the students; and some may have a different philosophy about the role of the teacher and how children learn best. A list of related problems would be generated by the new role of the principal.

**Compatibility of the Model with District Practices.** Although the extensive-requirements model may not be in serious conflict with the standards, guidelines, and practices of a smaller district, it would certainly conflict with those of a larger district. The larger district, which has the greatest need for reform and the greatest potential for savings with respect to patterns of implementation, typically has guidelines and standards in addition to those the state imposes. A simple rule for predicting conflict is that the more details the model addresses, the greater the potential for conflict. All details are controlled under the extensive-requirements model; therefore, the chances of the model being perfectly compatible with all district guidelines are effectively zero.

For instance, the district has its own provisions for professional development, which its guidelines articulate. At best, the content and practices of the district's professional development will not be greatly at odds with the content the model requires. However, the more time teachers spend in the course the district provides, the less time they have to learn the specific requirements of the extensive-requirements model. More probably, however, the district practices will conflict with many details of the model. Teachers engaged in such programs of professional development will, understandably, either become confused or assume that there are options when, according to the model, there are none. This is only one of a very large number of possible conflicts between model specifications and the specifications of the central administration—from the deployment of aides to the specific instructional sequences.

So, if the extensive-requirements model is to be fully implemented in a district, some sort of waiver from district guidelines is

needed. Without such a waiver, the model will certainly fail because the principal and teachers will be placed in the position of following both the district guidelines and incompatible model guidelines. Therefore, without a waiver that the district and the schools both honor, uniform implementation of the model is impossible.

### **Effective Models**

Models with extensive requirements can be divided into those guided by effectiveness and those guided by more-arbitrary criteria. The effective extensive-requirements models demand that all features of instruction that have been demonstrated to make a difference in the performance of students be controlled to produce a positive effect on performance. The model would therefore control the design of the curricular materials, the sequence of topics and lessons, how the material is delivered to the students, the type of student responses that occur during instruction, the reinforcement and correction practices teachers use, the grouping of students for instruction, the daily schedule of instruction, schoolwide disciplinary practices, the records of student mastery in the various instructional sequences, the procedures for using data to identify and respond to problems of student performance, and how the school celebrates academic success. The role of the principal would change to support the basic changes in teacher-student interaction.

The model the National Institute for Direct Instruction (NIFDI) sponsors has been shown to be effective in accelerating the performance of all students. The model requires procedures that are not used in most failed schools, and all teachers in the school are required to use these procedures. The model designates instructional programs that are explicit, systematic, direct, and precise. Coordinated schedules are established for the teaching of all subjects. Training focuses on how to use these programs—the presentation behavior and corrections that teachers are to exhibit in the classroom. NIFDI requires homogeneous grouping of students for instruction

and regrouping students as many as four times a year to ensure that students are placed appropriately in the program sequence. The model establishes precise expectations in the form of projections of lessons completed at mastery for every group of students in each subject area. Progress is monitored through direct observations, performance on in-program tests, records of the number of lessons each group completes each week, and weekly conference calls when the NIFDI consultant is not on site. The calls are used to identify possible problems (based on data reports) and to review the effects of remedies applied to previously identified problems.

### **Implementation Versus Stability**

A site may implement NIFDI in two years. Judgments about the quality of implementation are based primarily on classroom observations, which are used to answer a series of questions: Are the students placed properly in the instructional sequence? Is the teacher following the schedule and presenting material appropriately (wording, pacing, correction, praise, review)? Is the classroom set up appropriately? Has the teacher posted data regarding student performance? Are the data consistent with the projections for the various instructional groups and subjects?

The NIFDI model may be well *implemented* after two years, but it will not be *stabilized* until about the sixth year of implementation. The site is stabilized when all teachers are projected to teach about the same instructional sequence next year that they taught during the current year. It takes five or six years for this to occur in Grade 5 of a K–5 school.

During each of the preceding years, the material that is taught changes because the performance of continuing students entering every grade increases annually. During the first year, the fifth-grade teacher may have very few students placed in the fifth level of the reading program, many in the third level, and some in a corrective-reading sequence. In the second year of implementation, most of the incoming fifth-graders will be a year ahead of the previous year's students. There will be a larger number in the fourth and fifth levels of the program. By the sixth year, no continuing students (those who

started the program in kindergarten) will be in corrective reading, and very few will be below the sixth level of the program. Many will be out of the sixth level and in the final step of the sequence—studying history from a seventh-grade text.

### **Misinterpreted Data as an Impediment**

The misinterpretation of performance data can seriously impede scale-up of NIFDI or any other model that requires a substantial amount of time for implementing the model and stabilize a site. As mentioned above, fully stabilizing an effective site requires about six years in a K–5 school—the time it takes for the first students to pass through the program sequence from kindergarten through Grade 5. If the model is evaluated on the basis of the performance of fifth-graders after two years of implementation, the gains will not be impressive because the fifth-graders would have gone only through Grades 4 and 5 in the model, not kindergarten through Grade 5. The principal, teachers, or central administrators who observe the trends may conclude that the program works in the lower grades but not in the upper grades. They may choose to maintain the model in the primary grades but use other material and approaches for the upper grades. By modifying the sequence, they will likely still show some gains in student performance, but their approach would not produce the results that they would have produced by following the instructional sequence of the effective model. The only way the district can discover this relationship is to fully implement the model for five or six years.

The absolute number of sites implemented could be used to greatly distort data on site effectiveness. If a model developer is able to identify a dozen or more successful schools that use a particular model, the data may be compelling to a district, even if the model has a less-than-chance potential of achieving such gains. Let us say that three of five schools that use an extensive-requirements model are successful and that one of 40 schools that use the minimal requirements model is successful. If the extensive-requirements model has been implemented in 30 sites, 18 would be successful. In contrast, if the minimal-requirements model has been implemented in

1,000 sites, 25 would be successful. If the ratio of successful sites to total sites is not considered, the minimal-requirements model would appear to be more successful than the extensive-requirements model. It would certainly be able to produce more data on effectiveness. However, its apparent effectiveness is simply an indirect function of its ease of implementation and potential for adoption.

### **Fidelity of Implementation and Potential for Scaling Up**

Acceptance of a model that has the potential to be uniformly effective could be increased and its potential for rapid scaling up could be improved by lowering its standards in several different ways. For instance, more negotiation about the curricula and procedures could be allowed. Partial implementations that involve only some of the components of the model could be permitted, or it could use trainers who are not fully trained.

Although any of these approaches would make implementation easier, the data they generate would not provide realistic information to decision-makers about how to turn failed schools around. These approaches are problematic in at least three ways. First, schools in which the model is poorly implemented do not provide great benefits to the children and teachers. The implemented model would create some improvement (with great variability from one site to the next). However, the typical school would not serve all students and could not, therefore, provide a model for effective practices. Second, schools in which the model is poorly implemented do not present any compelling reason for a school to follow the tenets of the model. If the gains are modest, they could be achieved by a number of models. Third, these schools do not provide the district with data about what can be done. The rational district would have no inducement to make the changes in district policy necessary to follow the model closely unless it was evident that great gains are achieved only by schools that follow the model closely and that adhering to the model is manageable.

## **Patterns of Scaling Up and Their Relative Efficiency**

As discussed above, the most efficient method of scaling up an extensive-requirements model is through model clusters, which allow schools to maintain a high degree of fidelity to the model at lower overall costs.

### **Model Clusters**

Implementing an extensive-requirements model in clusters of schools is an efficient means of both scaling up and accommodating the high mobility that plagues high-poverty schools (25 percent per year and higher). When the first flight of students in an isolated model school passes through the program sequence from K–5, fifth-graders in the school will be performing as many as three grade levels above students from neighboring schools. Special provisions are required to accommodate incoming students because their performance levels are so much lower than those of continuing students. If the design is implemented in a geographic cluster of schools, however, a considerable proportion of students new to any of the schools will be transfers from other schools implementing the model. Placing these students in the instructional sequence is therefore simple.

Another reason for adopting the model-cluster pattern of scaling up is related to the fact that the extensive-requirements model is constrained by the relatively large amount of time needed to train trainers. With the model-cluster pattern, sites in which the model is fully implemented serve as training and dissemination centers. Teachers and on-site coordinators from schools in which the model is to be implemented are assigned to the model school within the district. This practice simplifies training and also provides teachers from schools that are new to the model with demonstrations of how classrooms are organized and how the program functions. Providing instruction through training and dissemination centers requires only about a quarter the time required to implement the model in isolated schools.

To increase efficiency, the model would be implemented in new schools largely through local coaches with the assistance of the model

developer. If the district does not permit a model school to function as a dissemination center, implementation of the model in other schools would proceed more slowly, simply because teachers and principals from these schools would not have people to model how the program was supposed to work. They would have to learn from negative examples instead of positive ones.

With advanced training of local coaches, the district could become virtually self-sufficient in training new teachers and implementing the model in new schools. The district would have to recruit a project director, who would assume management functions previously provided by the model sponsor, and institutionalize the positions of coach and project director. During this period, the sponsor's role would be reduced to overseeing the implementation effort and identifying problems of fidelity in implementation.

### **Difficulties in Establishing Fully Implemented Model Schools**

The time required to establish fully implemented and stabilized schools can create problems. The reason is that the gains in student performance tend to show up first in the grades in which students are closer to the norm. On such measures as reading level, the farthest a student could be below the norm at the end of Grade 1 is about one year. The farthest a fifth-grader could be behind is five years. If the students have historically progressed at the rate of 0.6 year gain per school year, they will perform on average around the third-grade level by the time they complete Grade 5. Given that attaining growth of much more than one year for each year students are in an effective program is unlikely, the pattern of improvement will show up first in kindergarten and Grade 1, then Grade 2, then Grade 3, and so forth. At the end of the first year, students in kindergarten and Grade 1 may show substantial gains over previous performance. The next year, the pattern will change because the first-graders, who have gained one year's worth of skills, will start on the Grade 2 level of the program.

The domino effect continues, a grade at a time. However, observers often misinterpret the progress by assuming that the difficulty of stabilizing the various grades is equal. After the second year, they may observe that children in the beginning grades have

improved greatly, but fourth- and fifth-graders are still behind even though they had been in the program for two years. Observers may conclude that the program does not work well with students in Grades 3 and above and may modify the program in a way that dilutes its effects. They may retain the instructional sequence for the primary grades and do something else in Grades 3–5. The remedy for this misconception about the data is to provide the site with information both about the anticipated trends and the reasons they will emerge.

A related impediment to implementing the model fully is what may be called “premature elucidation.” Often, school staff members and administrators see students perform better than they had historically after implementing the model for one or two years. Administrators identify some of the features of the design and conclude that they understand how the gains were achieved. They may modify or eliminate aspects of the design they do not consider crucial for student success.

If student performance remains the same or increases, these administrators will feel justified in having modified the design. The increase in student performance may be attributed to the new regime, even though it is actually the result of the model’s implementation with earlier cohorts of students. In any case, staff members and administrators do not receive information on how inappropriate the modifications are because they do not receive information on how well the students would have performed if the modification had not been introduced.

### **Special Challenges of Large School Districts**

One of the most serious problems of implementing an effective model in the larger district is to carry out the implementation in such a way that the model is not viewed as a foreign body at odds with the district. The potential for such a perception originates in the conflict between procedures that are effective with students and procedures the district guidelines and standards mandate. If waivers are not pro-



vided and honored, the schools in which the model is implemented will be subject to the district guidelines. At least some guidelines will be inconsistent with the specific provisions of an effective model. For instance, a district's rules for behavior management may be reasonable but very general. The rules for behavior management that the teacher follows in an effective model may be very specific. Implementing the district guidelines or presenting general training to the teachers would then be a relative waste of time because the specific training the teacher receives covers everything the general presentation covers but references behavior management procedures to the specific details of the various instructional programs and activities of the model.

If the school follows the district guidelines for test preparation, professional development, and the scheduling of curricular events, the schools will not perform as well as they would have if they had followed the specific guidelines of an effective model. This assertion is based on the fact that the district guidelines have never been demonstrated to be highly effective, but the guidelines of the model have. If a school ignores the district guidelines and is highly successful, it will not be seen as a product of the design's procedures. Rather, it will be viewed as a joint product of the design and the district test-preparation program, professional development, and curricular requirements. Schools that follow this joint formula will tend to fail.

These problems are serious, both because they preempt the district from discovering the model's potential and because data from the school constrained by the district guidelines will show only what the model does when constrained by those guidelines. The district's position is, in the broadest sense, paradoxical. In the past, the district has failed. Even if it has adopted new guidelines, it has no basis for undaunted confidence that they will work. Various guidelines adopted in the past have never worked. Yet, the district often remains staunch in requiring models to follow the current guidelines. From the standpoint of simple probability, any model capable of achieving great improvements in teacher and student performance will tend to fail if its implementation is attenuated by the district requirements.

The most productive way for an effective model to work in a district, therefore, is for implementers to receive a waiver from the district guidelines. The waiver indicates that the design implementer will be responsible for teaching all the skills needed for the students to perform well on the achievement tests, including the state test. In the political world of school districts, such waivers are difficult to obtain because they seem to suggest that the model implementers have no respect for the district's guidelines. However, the district needs to learn at least one effective formula for achieving accelerated performance of at-risk students. The best way to obtain information about what results are achievable and what structures and support are needed to achieve them is to implement effective models fully, then to evaluate them.

### **Specific Problems with District Guidelines**

District guidelines have three characteristics that may create problems in implementing effective school reform models: Some function as a curriculum, some support laxity, and some tend to require work on material that is of only peripheral importance to accelerating student performance.

**Guidelines as Curriculum.** Guidelines function as a curriculum when they specify a pedagogical process rather than learning outcomes that are reasonable for a particular grade and subject. A process is implied by every standard or guideline that requires schools to teach something before it would be taught in an effective program sequence. Such processes may override sensible instructional sequences. For instance, if the district (or state) guidelines call for teaching the fractions  $1/2$ ,  $1/3$ , and  $1/4$  in kindergarten, the guidelines are not serving as standards but as a curriculum. The teachers are now required to follow this curriculum even though it does not represent a sensible way to introduce fractions or a sensible time to do so. The guidelines do not indicate an outcome that is important for going into Grade 1 or even Grade 2, yet they are very specific about what is to be taught. The guidelines fail to recognize not only that teaching fractions in kindergarten is unwise but also that this

sequence of fractions will probably reinforce misconceptions about what fractions are and how they are related to the counting numbers.

There are many other examples of guidelines that function as curricula. For instance, if the guidelines require students to work on a particular type of word problem in fourth-grade math, even though it is doubtful that they have the math skills necessary to solve such problems, teachers must now somehow teach these skills. The idiom of “writing as a process” is reasonable in some ways, but the steps the district may require are certainly not the only set of steps that will lead to good writing. Furthermore, not all the writing the students do should be of the form that involves note-taking, first draft, revision, and publishing. Successful programs that emphasize students’ writing more and writing in a way that yields better first-draft material should not be forced into the Procrustean “writing process” mold.

**Lax Standards.** The second type of failure is created by guidelines that are too lax in that they do not require performance on a skill until long after it would have been taught in a reasonable instructional sequence. The curricular sequence is affected far less by lax specifications than by guidelines that act as a curriculum, but the credibility of the sequence is still challenged. Lax standards provide justification to teachers for not following the specifications of a validated sequence. For instance, districts may adopt the guideline, “Read by Grade 3.” NIFDI has consistently demonstrated that if a reading sequence is properly implemented in kindergarten, virtually all at-risk students with the exception of the profoundly retarded and the very frequently absent will read by the end of the year. No program that purports to be a model of reform should have a standard less demanding than “Read by Grade 1.”

**Guidelines That Stress Peripheral Skills.** Guidelines that stress peripheral skills create two problems. First, because they do not test key skills, they suggest that these skills are not important. Second, they test skills of questionable value, thereby implying that these skills are important. For instance, math tests—both standardized achievement tests and district- or state-created tests—tend not to test math skills that are absolute prerequisites for higher math and, instead, tend to test trivial skills and applications.

For example, one of the skills essential for higher math is facility at writing and rewriting equations. This skill is not included in many tests. Instead, tests typically present problem types that students have not learned how to express as equations. Much of what is tested is inconsequential from the standpoint of mathematics. Blueprints, graphs, and virtually anything that has numbers are treated as legitimate math items. Certainly, students should learn this material, but most of it is not really legitimate math content and should not replace legitimate math content.

The main problem with guidelines that stress peripheral skills is that teachers become reluctant to follow an effective program because much of what is taught in the program is not tested. Understandably, the teachers are likely to see the program, not the guidelines, as problematic.

### **Educating the District**

One of the great difficulties in implementing an effective reform design is to educate district-level administrators on specific details of the model. Often, there is no clear channel of authority within the larger district, which means that it is difficult to identify the administrators who should become informed about the model. Attempts to communicate with the school board prove abortive in larger districts because the board does not have provisions for directing the administration in curricular or instructional matters. There is no decision-maker within the district who has clear responsibility for the school in which the model is being implemented. It may be that the superintendent in charge of elementary education has more power than the superintendent in charge of reading instruction, the regional superintendent, or the head of the office of accountability. In practice, however, the responsibilities are not clearly delineated. The school may be bombarded by input from all of these administrators, plus, possibly, the head of English as a Second Language instruction, the Title 1 coordinator, the school's probation officer, the state's probation officer, and the director of special education.

There is little likelihood that model implementers can communicate effectively with the sources of all these inputs and inform those

who provide directives to the school about how the program is designed and why they should subordinate their guidelines to those of the model. Even if the district agrees to a waiver, however, it is often not honored, largely because it is not part of the district's organizational structure or patterns of interaction with schools. The result is continual input from the district that conflicts with the requirements of the model.

For example, NIFDI dropped all its schools in one large urban district because there was no probability of fully implementing the model in them. NIFDI had an agreement that specified that the district would not introduce practices that conflicted with the model's requirements; however, the conflicts were legion. The district had adopted the policy (which is now thankfully rescinded) that students were to be instructed only in "grade level" material. If the children were in Grade 3, all instruction had to be provided with third-grade material, even though a large percentage of the students transferring into Grade 3 read on the first-grade level. In one school that had a fairly large population of non-English-speaking children, the ESL director prevented NIFDI from installing a sensible English language program in kindergarten and Grade 1 and insisted on using a program that basically taught Spanish. Neither the principal nor the teachers in any of the schools understood that they were to follow the requirements of the NIFDI program. They responded to the NIFDI implementation managers as sources of possible ideas, not as trainers. For them, the true orders came through the regional educational officer and other district administrators.

Although the results that were achieved in the schools NIFDI worked with were far better than the average of the at-risk schools in the district, the district was not learning anything significant about what could be achieved with a coordinated effort. NIFDI was providing a service for some of the students and teachers in that locale, but the model would have been far more productive in a place with the potential to support the implementation.

In another large district, NIFDI found itself at odds with the district about the kind of preparation needed for the state test (which is horrible and fortunately is being rescinded). NIFDI's position was

that the students would be farther ahead if the teachers did not follow the district guidelines about providing extensive “test prep” instruction (daily, from October through March) but used the time to move students farther through the NIFDI curricular sequences. Because of this and other conflicts, the NIFDI model was implemented in varying degrees of fidelity in the schools—from one that basically followed none of the district guidelines about curricula or training to those that tried to follow all of them. The performance data in the fifth and sixth year of the implementation confirmed the correlation between following the model and improving student performance. The school that followed the NIFDI guidelines most faithfully was originally one of the very lowest performing schools in the district. The mean reading scores of the school’s fifth-graders on the 1998 reading portion of the Comprehensive Test of Basic Skills (CTBS) was at the 14th percentile. Three years later, the mean Grade 5 reading score was at the 67th percentile, one of the highest scores in the district. The school also had a larger percentage of fifth-graders passing the state test than any of the other NIFDI sites—all of which started higher than this school.

The tragedy is that the district does not view this school as a model of what is possible. The district has made little attempt to learn from this implementation. Although it is a supreme exemplar of what is achievable and although the district has never seen a school achieve a turnaround of this magnitude, the school serves more as a political thorn than as an indicator of the result of implementing effective guidelines for accelerating at-risk students.

### **Nested Responsibilities**

The problem of diverse and often inconsistent inputs to the school implies that the larger district needs an organizational overhaul if it is to support and scale up effective models in a way that does not seriously compromise their integrity. The organizational design would simply funnel all inputs to the school through one central administrator. This administrator could be in charge of perhaps six schools. The administrator’s job would be to facilitate the implementation of the model by responding to problems of implementation in a timely

manner and ensuring that all inputs are consistent with the model but not in violation of relevant board or union requirements. The administrator would have the arbitrament of reassigning teachers and aides, deploying and training full-time substitutes, providing time for preservice and in-service training, coordinating efforts of the various departments of accountability, elementary education, and so on, as well as overriding or adjusting any requirements from these departments that are inconsistent with the requirements of the model and the board's agreement with the model.

### **Timely Remedies**

The slow pace of responses to problems is pandemic in larger districts. This delay is particularly devastating during the first two years of the implementation because the patterns established during these years tend to persist. One of the more predictable problems is the teacher who receives training but does not follow classroom assignments. In October, the problem may be identified and noted as one that, if not solved, will result in the children being far below the end-of-year performance level projected for them. According to our experiences in working with more than 20 large school districts, the problem typically will not be addressed during the current school year without a central administrator who has the power to effect some form of solution, and the chances of it being addressed at the beginning of the next school year are less than 50 percent.

These and similar problems are not easily solved on the school level, even if the principal tries to solve them in a timely way. For example, the principal's only practical solution to high absenteeism of teachers on Monday is to regroup students and do the best that can be done. The typical solution for the teacher who is not following assignments is to have the coordinator or another teacher teach her groups, creating a domino effect that reduces the capacity of the school and the coordinator to respond to other problems.

The central administrator could solve this problem in far more direct and effective ways. These remedies would be timely, which is particularly important for students who are already seriously behind their peers.

## **Consistent Inputs to Schools**

All district inputs to the school would be screened by the central administrator. Those that are clearly inconsistent with the model would be vetoed; those of questionable compatibility would be discussed with the model sponsor. Any requirements of the model that seem to be unreasonable would be negotiated. The principal would have regularly scheduled meetings or conference calls with the central administrator and model sponsor to discuss implementation problems and progress.

The central administration that had provisions for implementing reform models with fidelity would have ongoing data about the relationship of the schools' compliance with the model and teacher-student performance. This information is essential for districts to make informed decisions about which models to support, what type of support is needed, and the benefits and relative cost of their full implementation.

## **Conclusions**

An extensive-requirements model that has the potential to produce uniform acceleration of students is something of an island of “extropy” surrounded by entropic forces that compromise its implementation. If it is successful, it differs from what districts and schools do now. It is not global but is specific enough to ensure that teachers know exactly what they are expected to do and that what they do works with the students. The success of the model depends on a coordinated effort, which means that unless all teachers play their roles, the potential gains will not be realized. If even a third of the teachers in K–2 “do their own thing,” for instance, the implementation will show only modest gains in student performance. The model achieves acceleration not through magic but through careful control of all the school-related variables that affect student performance. The use of time is maximized to make the school environment effective—not nervous, impatient, or hectic, just very effective.



Teachers who participate in this model must be trained, and often retrained, in how to present material and how to correct and interact with students. This training probably should have occurred while the teachers were in college, but, in most cases, it did not. So if it is to occur, the model must provide for it as part of the implementation process. In the same way that teachers must be trained, trainers must first be effective teachers and then receive training on how to work effectively with teachers.

In part, the impediments that an effective model encounters result because the model is different, because the model requires hard work, and because it has strict performance standards for teachers. The most serious problem, however, is the resistance of larger school districts to approaching the adoption of the model as an information-gathering enterprise. The district is failing, which provides strong evidence that it does not have either the structures or procedures necessary to be effective with at-risk students. A model that has the potential to accelerate the performance of all students will provide the district with the information it needs on at least one way to be effective. For the district to receive this information, however, the district must recognize the requirements of the model and honor them. Unless the district assumes that, under the model, school performance will be worse than it is currently, the district should have no trouble with the idea of waiving requirements that are inconsistent with the model and implementing it thoroughly enough to evaluate it. If the evaluation reveals that the gains are only slight or are not correlated with the degree to which the model was implemented in the schools, the district could scrap or modify the model. If the gains are unparalleled, however, the most compelling implication is that the district should make the changes necessary to expand the model.

Within any district that views a model developer not as a vendor but as a partner who has a system that works, the problems of increasing the scale of an intervention in a reasonable period without compromising the model are solvable. The predictable decay that occurs in well-implemented schools that are not closely monitored shows that no systematic solutions will occur until districts recognize what is involved in achieving uniform acceleration of students and

therefore what must be done to institutionalize the practices so that the system has built-in monitoring and remedies for problems.

In the absence of these provisions, any model that has the potential to accelerate student performance, particularly in larger cities, will affect only the teachers and students in the schools where the model has been fully implemented. This effect will probably not continue much beyond the period the sponsor works with the schools. In the meantime, the district will continue to specify and implement new guidelines, standards, and rules that may produce modest gains over the current status but will not begin to show what could be achieved with a model that carefully controls the details relevant to teacher-student success.

Ultimately, the most serious impediment to scaling up effective models is the school district's structures and practices. Scaling up would be greatly simplified if districts were organized so they could implement effective models with fidelity. With a responsive district, the model could be implemented in school clusters, and effective training formats could be applied. Most important, effective practices could be identified and institutionalized so the district would gain the capability to train its teachers and closely monitor its schools to ensure that schools not only achieve high levels of student performance but also maintain them. Instead of issuing new standards and guidelines based, for example, on the alleged performance of students in New Zealand (whole language) or England (the open classroom), the district would have ongoing data about what works in its schools and would continue to implement demonstrably effective practices so long as the district is charged with the responsibility of educating at-risk students.

## References

- Adams, G., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis*. Baltimore: Center for Research on the Education of Students Placed At Risk (CRESPAR), Technical Report 59.
- Carlson, C. D., Francis, D. J., Latif, L., Priebe, S., & Ferguson, C. (2002). *RITE program external evaluation: Executive summary 2001–2002*. Houston: The Texas Institute for Measurement, Evaluation, and Statistics.
- Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The Direct Instruction Follow-Through Model: Design and outcomes. *Education and Treatment of Children, 11* (4), 303–317.
- Gersten, R., & Keating, T. (1987). Long-term benefits from Direct Instruction, *Educational Leadership, 44* (6), 28–31.
- Gersten, R., Keating, T., & Becker, W. (1988). Continued impact of the Direct Instruction model: Longitudinal studies of Follow-Through students. *Education and Treatment of Children, 11* (4), 318–327.
- Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model* (Vol. IV-A: An evaluation of Follow-Through). Cambridge, MA: Abt Associates.