Chapter Seven

THE FUTURE OF WHOLE-SCHOOL DESIGNS: CONCLUSIONS, OBSERVATIONS, AND POLICY IMPLICATIONS

Mark Berends, Susan Bodilly, Sheila Nataraj Kirby

The NAS effort, as we mentioned earlier, offered an unprecedented opportunity to study and understand a dramatic attempt at whole-school reform from its inception—one based on an experimental approach of research and development, demonstration, and scale-up. Analyzing the initiative over this past decade has offered a broad perspective on the issues facing different parties to the reform effort: founders, developers, district administrators, principals, and teachers. Perhaps more importantly, RAND's findings from monitoring the NAS effort over this past decade offer important and sobering lessons for federal, state, and local policymakers attempting to improve failing schools through comprehensive school reform. This final chapter provides a retrospective on the following:

- The unique nature of the RAND approach to evaluating the NAS initiative, its strengths and weaknesses;
- The contributions of NAS to educational reform;
- The implications of the RAND analyses to the external change agent theory of action; and
- Implications for the current federal CSRD policy.

UNIQUENESS OF THE RAND APPROACH

While NAS has been a unique effort, the analyses performed by RAND concerning NAS have also been unique. The scope of the NAS effort was remarkable: capitalizing seven teams; pushing for a scale-up effort in ten districts; and undertaking a scale-up effort in hun-

dreds of schools. The RAND analysis had just as remarkable a scope. It covered the evolution of NAS and its teams from inception to 1998. It tracked implementation and performance in over one hundred schools from 1996 through 1999. It combined an action-oriented approach with formative assessment methods conditioned on the developmental nature of the intervention. For that reason, the RAND analyses focused heavily on the development of designs and their implementation strategies. These two issues are overlooked in more-summative research designs. This approach enabled NAS and RAND to draw strong connections between the conditions of implementation and outcomes.

This uniqueness had both strengths and weaknesses. Its strengths were that it provided in-depth information on implementation and systemic issues that few other studies of reforms have provided, and it uniquely identified both the collage of actors and activities that needed to be successfully meshed for the effort to prove successful and the important conditions under which the innovations could flourish.

On the other hand, it did not provide as clear and compelling evidence on outcomes as one might have desired, especially if the sponsors were interested in significant improvements for selected groups of students. Given the developmental nature of the intervention, the measures and controls were not and could not be geared for this. Furthermore, the teams themselves often did not provide the proof of their theories of learning that would have complemented the RAND efforts and strengthened the entire analytical impact. However, the framework and methods were strong enough to show that dramatic improvements expected by NAS and the partnering districts did not exist for most schools.

The RAND approach worked well under these very special circumstances: a developmental intervention; a client needing information for improvement quickly; and the stipulation that the results had to be dramatic to continue with the effort. However, sponsors interested in more-specific and careful measures of more fully developed interventions would not benefit by this approach.

THE CONTRIBUTION OF NEW AMERICAN SCHOOLS

NAS accomplished several of the goals it had set for itself and in the process made several important contributions to educational reform that need to be kept in mind.

NAS funding and leadership led to the deliberate development of functioning design teams. When NAS started, few whole-school designs existed. Judging from the initial proposals, there were few organizations capable of design development or of thinking in terms of whole-school approaches. NAS funding eventually took seven designs from ideas on paper to firmly developed and functioning teams whose designs have been adopted in multiple sites in the educational marketplace. In essence it proved that teams could be deliberately created and developed over time.

NAS showed that initially dependent external change agents could be moved toward self-sufficiency over time. NAS undertook a very systematic venture capital approach. It weeded out teams it thought were not moving toward strong self-sufficiency or showing the results it wanted, while it provided capacity-building funds for those that were moving in its preferred direction. It provided assistance to teams in their development to allow them to become self-sustaining. Perhaps most notably, NAS encouraged federal funding, in the form of advocating for the CSRD program, to allow for growth in the market for the teams. In the end, NAS still provides some funding through loans for capacity building, but the remaining design teams otherwise operate independently in the marketplace of school reforms.

NAS explicitly sought scale-up of the reform initiative. Foundations and others have sought to create self-sustaining programs. Few have been able to do so, but some have been successful. NAS went a step further and deliberately pushed toward a scale-up strategy from the very beginning. NAS can take credit for deliberately spreading, or scaling up, a variety of designs in many different school settings. The creation of the CSRD funding is just one manner in which NAS still promotes the concept of scale-up of design-based assistance for whole-school reform.

NAS explicitly made analysis and good consumer education a part of its efforts. A major purpose behind the RAND analyses was to of-

fer critical and timely analysis of the NAS operations. NAS has always been very supportive of this effort and an avid consumer of the RAND information, using it to help identify problems and solutions to improve the initiative. From the beginning, NAS has emphasized the importance of quality assurance among its teams. It continues to promote this idea through the *Guidelines for Ensuring the Quality of National Design-Based Assistance Providers* and by creating and spinning off the Education Quality Institute.

Our review of the NAS experiences indicated that this deliberate effort did succeed in some important ways, and the approach of providing venture capital with specific goals could be used as a policy instrument in the future when innovative approaches and new actors are desired. In addition, NAS actions as a change agent have significantly influenced policy in its areas of interest.

IMPLICATIONS FOR THE EXTERNAL CHANGE AGENT THEORY OF ACTION

We began our discussion by indicating the RAND's analyses focused on the theory of action inherent in the NAS initiative, not on the efficacy of each design. We return here to that theory as a way to draw conclusions about the NAS initiative. RAND findings provide mixed evidence to support several hypotheses underpinning NAS's theory of change:

- The initial NAS hypothesis, that by adopting a whole-school design a school could improve its performance, was largely unproven. We found specific positive examples of schoolwide implementation and improvement under certain conditions; however, negative examples were found under more-common conditions. Our general findings were of weak implementation and lack of strong improvements in school performance. The RAND analyses provide neither clear support for nor evidence against the contention of some that whole-school design approaches are superior to more programmatic approaches to school reform (Slavin, 1997a, 1997b, 1999, 2000; Pogrow, 1998, 2000a, 2000b; Slavin and Madden, 2000; Fashola and Slavin, 1998).

- NAS's and RAND's hypothesis that designs alone are not helpful to schools and that schools need assistance in implementation was proven correct. Teachers and school administrators clearly reported higher levels of implementation associated with strong assistance from design teams. But, just as importantly and consistent with the implementation literature, conditions at the schools and within the districts and the manner of selection also proved important to implementation and outcomes.
- The scale-up hypothesis that a district that converted 30 percent of its schools using whole-school approaches would become high-performing and not revert to unproductive practices was disproved. Districts, such as Memphis, reverted back to their former status quickly with changes in administrations.
- The scale-up hypothesis that a district needs to provide a supportive environment was dramatically proven by the negative case of San Antonio. Without a supportive environment the designs did not flourish.

In general, we conclude that the theory of action was largely underdeveloped and underspecified. The causal chain of events leading to strong implementation and outcomes has proven to be far more complex than that originally considered by NAS and one that remained largely outside of its control and influence. This finding is in keeping with the literature on implementation indicating the complexity of the change process.

Based on our experience with NAS, we offer the following implications for future efforts to bring about whole-school reforms through external agents.

Externally developed education reform interventions cannot be "break the mold" and still be marketable and implementable in current district and school contexts. NAS attempted to have both "break the mold" designs and designs that would appeal and be implemented nationally. It faced and still faces a fundamental market issue. The evidence of our evolution analysis and the implementation analyses all point to the fact that schools did not have a ready place for these designs. Schools were not by and large fertile ground for "break the mold" ideas, often because of a lack of capacity or local, state, or district regulations. Rather, the designs had to change to

be suitable to school conditions or simply not be implemented. Design team calls for significant school autonomy over budget, staffing, curriculum, instruction, and assessments often did not fit into the institutional infrastructure that schools faced. Under these conditions the designs often settled for approaches that called for marginal improvements over time. In order for the design to be well-implemented, the district and school contexts have to change to allow for "break the mold" school-level ideas to flourish.

External interventions need to address systemic issues that can hinder implementation. The relatively weak implementation of the designs in scale-up was associated with several systemic factors: lack of teacher capacity to undertake the designs, especially in terms of time and subject area expertise; lack of principal leadership; and an incoherent district infrastructure that did not match the needs of design implementation. Improved district support appears difficult to obtain, but perhaps feasible with a significant resource outlay and strong relationship between a NAS-like organization and a district (not just the superintendent). The requirements for teacher capacity and principal leadership for design implementation appear more problematic. It is those very schools lacking in these qualities that theoretically would most benefit from external assistance interventions. This implies the design concept did not focus on the important dimensions of school improvement when attempting to increase school effectiveness. Greater attention to building basic teacher capacity and effective principal leadership for transition should be the focus of reform, at least for low-performing, high-poverty schools.

A rush to scale-up when interventions are not completely developed weakens results. NAS designs and teams were not ready to scale up when NAS called for this move in 1995. NAS was not ready for this scale-up either. It had not fully developed its concepts of school matching, district partnerships, or a supportive district environment to the point where it could ensure its designs would thrive in a scale-up activity. Many of the problems associated with the scale-up phase are attributable to pushing toward full-scale production before all the kinks in the product were worked out. However, these problems are likely to persist partly because developers are under financial pressures to scale up their interventions before they are thoroughly evaluated and partly because districts and schools are under severe political pressure to adopt external solutions—whether

proven or not—as a means of addressing the lackluster performance of their students. Venture firms like NAS must weigh the benefits of waiting until development is complete against the costs of waiting to enter the field with a new product when interest is high. The consequences of not waiting should be taken seriously. Failures in implementation due to a rush to get into schools result in a continuing weakening of the trust between teachers and administrators and between parents and the education community.

A key component of successful implementation is consistent, clear, and frequent communication and assistance between design developers and schools, particularly teachers. Implementation tended to be very low in schools where teachers reported they did not understand the design, see the design team members often, or receive strong assistance. Case studies pointed to the importance of the selection process in establishing firm positive relationships between the external agent and the school staff; in addition, they highlighted the importance of communicating with and involving all teachers in the school, not just the "leads." The external agents struggled to provide high levels of assistance, but could not always do so given the leap in capacity required from demonstration in a few, often nearby, sites to scale-up in many schools across the country. A reasonable inference from our research is that a strong, trusting relationship between a school and an external agent is a prerequisite for strong implementation of complex interventions that require significant changes in behavior. If funders and developers expect teachers to change behavior significantly, then they need to invest considerable time and effort to build trusting relationships with teachers.

Monitoring site progress, self-evaluation, and reflection are necessary if external developers are to be successful and to improve their offerings over time. Our work throughout the development and demonstration phases indicated that most of the teams had not created the feedback loop or data needed to further develop their designs and offer meaningful support packages to the schools. This affected their ability to produce the desired results. In later phases, most did not collect adequate information about their implementing sites to allow for proper support or even to hazard a guess as to implementation levels in schools. This ran counter to the emphasis placed by NAS on quality assurance and its continuous calls for self-evaluation by the design teams. In part, this is a resource issue—the

push to scale-up left the developers with few resources for evaluation. In part, it is a priority issue—developers want to spend money on development of the ideas they are committed to, oftentimes whether or not they are effective, and may see evaluation as less important or too expensive. But, unless systems for tracking progress in schools and understanding school-level concerns are created and used for improving the external intervention, then the effort cannot succeed over the long term. This capacity must be deliberately built into the development effort and continuously maintained.

The typical outcome measures used in public accountability systems provide a very limited measure of student and school performance. Years of evaluations indicate that the best way to measure whether an intervention is having an effect is to measure variables most closely associated with the interventions. This truism would lead evaluations away from using district and state test score data toward a richer set of assessments and indicators for whole school reform. However, the conditions in the scale-up phase worked against this. First, the external agents had not developed convincing indicators of progress; in fact, they had to be asked to develop benchmarks for progress by the districts because indicators of progress did not always exist. Second, districts insisted their tests be used as the sole indicator and NAS promised dramatic improvements in this single indicator. In short, in the developmental phases of an intervention, the assessment instruments needed to adequately measure progress do not exist. The assessment measures that do exist—district-mandated tests—do not adequately measure the impact of innovative approaches.

This tension will be a constant hindrance to understanding the impact of innovative approaches unless alternative indicators and assessments are developed in ways that are well aligned with what the reforms are trying to do. Few reforms will show strong results when they are geared toward improving students' complex thinking and mastery of difficult subject matter, but are measured by simplistic tests. We will not be measuring what is important, but measuring what is easy to measure. The high-stakes testing regimes currently in vogue and the overwhelming emphasis given to improved test scores on state- or district-mandated tests as *the* measure of improvement do not bode well for many innovative reform efforts.

IMPLICATIONS FOR CURRENT POLICY: A CAUTIONARY NOTE

Currently, many schools throughout the country are attempting whole-school reform requiring significant changes in teacher and administrator behaviors using the federal funding provided by such programs as Title I and the CSRD program. RAND's program of studies of NAS has identified the conditions needed to make these efforts successful including: teacher support and sense of teacher efficacy; strong and specific principal leadership abilities; clear communication and ongoing assistance on the part of design developers; and stable leadership, resources, and support from the district.

The RAND analyses indicate these conditions are not common in the districts and schools undertaking CSRD—schools with similar characteristics to those NAS served in the scale-up phase. Because the target of the federal Title I and CSRD funds is primarily high-poverty schools, schools most likely to be affected by the CSRD program are also schools that are most likely to face very fragmented and conflicting environments, difficult and changing political currents, new accountability systems, entrenched unions, serious lack of slack resources in terms of teacher time, and demoralized teachers given the fluctuating reform agenda and the difficult task of improving student performance under these types of conditions (for a description of CSRD schools see Kirby et al., in review). These schools will face many obstacles during implementation of whole-school designs, and because of this, whole-school designs will face continuing challenges in significantly raising the achievement of all students.

Given this, federal and state policymakers need to think critically about their current stance of simultaneously promoting high-stakes testing; the implementation of comprehensive school reforms that promote innovative curriculum and instructional strategies; and the implementation of multiple other concurrent reforms. This is especially the case when confronting reduced state and local budgets during a time of retrenchment.

The implementation of high-stakes testing regimes precludes the adoption of rich and varied curricula that challenge students and motivate them toward more in-depth learning experiences. It cer-

tainly prevents adoption of such curricula when other more basic skills reforms are mandated on top of the design-based curriculum. High-stakes tests become a two-edged sword in this environment. On the one hand, high-stakes tests motivate schools to increase performance and often to seek out new curriculum and instructional strategies associated with comprehensive school reforms. On the other hand, those very same tests provide disincentives to adopt richer, more in-depth curriculum—even when mandated.

Concurrently, these same districts are facing new and growing pressures to see the performance of their lowest achieving schools increase substantially, and these schools are frequently high-poverty, high-minority schools that receive Title I funds. For example, some of the key Title I provisions involve states establishing rigorous standards for what students should know (content standards); establishing performance standards for how well students should know the content; and developing assessments to measure school and/or student progress toward these goals. By the spring of 2001, all states were required to have such assessments in place to comply with Title I policy, with additional testing and accountability requirements emerging from recent federal legislation.

If districts react to this pressure with past behaviors, they will likely promote the failure of whole-school reforms. In the past, districts have sought to increase accountability, while also mandating a series of reforms, without providing for the slack needed to implement them. For instance, some districts are mandating reading and math programs with specified professional development routines, increased teacher and principal accountability based on inappropriate test regimes, and further reductions in school-level budgets. Simultaneously, they encourage schools to adopt schoolwide models without much review of effectiveness or fit with the district policies or school needs. The result will be, as it was in the scale-up districts, continued fragmented, incoherent policies not supportive of whole-school interventions.

In short, we anticipate continuing conflicts between whole-school design or model adoption and district and school contexts as well as political pressures rushing schools and external assistance providers into partnerships that are not well thought through. If districts continue in this manner, the outcome will be neither short-term gains,

nor long-term success. Expectations regarding the ability of schools to make meaningful changes with the assistance of externally developed designs in this fragmented and unsupportive environment are not likely to be met. This may well lead policymakers to abandon what could be a promising vehicle for whole-school reform without having given it a chance.