
SUMMARY

New American Schools (NAS) was formed in 1991 to create and develop whole-school designs that would be adopted by schools throughout the country in order to improve student performance. It was established as a nonprofit and funded largely by private sector donations. NAS founders thought that in the past many reforms were “programmatically,” focused on a particular set of individuals in a school or a particular subject or grade level. They believed that adoption of multiple and unconnected approaches to address each area of schooling resulted in a fragmented education program, a balkanized school organization, and low performance by students.

NAS’s core premise was that all high-quality schools possess, de facto, a unifying design that allows all staff to function to the best of their abilities and that integrates research-based practices into a *coherent* and *mutually reinforcing* set of effective approaches to teaching and learning for the entire school. The best way to ensure that lower-performing schools adopted successful designs was to fund design teams to develop “break the mold” school designs that could be readily adopted by communities around the nation. After developing the design, teams would go on to implement their designs in schools throughout the country. This adoption would lead to NAS’s primary goal of improving the performance of students.

This whole-school approach to educational improvement was a dramatically different way of initiating and disseminating large-scale educational improvements. It was a unique combination of (1) private sector involvement using a venture capitalist approach; (2) the choice of whole-schools designs as a vehicle for reform; and (3) the ambitious goal of scale-up across the country.

The NAS experimental approach required careful development and demonstration of designs prior to moving to scale-up; therefore it had a phased approach including:

- Competition and selection phases (1992);
- Development phase of one year (1992–1993);
- Demonstration phase of two years (1993–1995 including the 1993–1994 and 1994–1995 school years); and
- Scale-up phase of three years (1995–1998 from the school year starting in 1995 to the one ending in 1998).

This approach to educational improvement offered an unprecedented opportunity to study and understand a unique attempt at school reform from its beginnings to its completion. Any educational reform must have two components, a theory of learning and a theory of action. Following Fullan (2001, p. 187), a *theory of learning* focuses on the assumptions about how students learn, instructional strategies, and performance. A *theory of action* focuses on the local context such as the conditions under which a design or external model will work. The theory of learning in the NAS initiative was embodied in the individual designs. Design teams were to be responsible for their own development of a theory of learning and for its evaluation. This left open the issue of whether NAS’s theory of action—at that point very unspecified—would be effective. To show important policy audiences that the NAS approach was efficacious, NAS would have to assess its theory of action as to whether it and the design teams could get schools to implement designs, and if and when schools did implement, whether this would lead to improved student performance.

RAND’S ANALYTIC TASKS AND PURPOSE OF THE STUDY

NAS approached RAND to assess and analyze its theory of action and to provide analytic support to its school reform efforts. This support took many forms, but primarily it was intended to document and analyze the conditions under which NAS made progress toward its goals of widespread use and implementation of its designs and improved student performance associated with that use. It included the following analytic tasks:

- Document the NAS efforts to assess its contributions to education reform;
- Describe the designs and analyze changes in them over time;
- Assess the level of implementation in design-based schools during the demonstration and scale-up phases;
- Identify factors that impede or encourage implementation in the demonstration and scale-up phases; and
- Measure whether the adoption of the designs resulted in the outcomes desired by NAS and its partnering districts in the scale-up phase.

These tasks were conducted over the first seven years of the NAS initiative from the demonstration phase through the scale-up phase, and the results are documented in a series of RAND reports, listed in the Preface.

The purposes of this book are to provide a retrospective look at the NAS initiative and the various RAND studies of that initiative; to draw together the findings from the diverse tasks undertaken on behalf of NAS; and to reflect on the lessons provided by this initiative for future school reform efforts. The document should be of interest to both policymakers and practitioners in K–12 education as it offers some important lessons learned about attempts to significantly improve student performance through the adoption of externally provided interventions.

A BRIEF HISTORY OF NAS FOCUSING ON ITS THEORY OF ACTION

To understand the RAND assessment of NAS, one must understand that NAS's theory of action and design teams' theories of learning were evolving through the entire period of the analysis. Therefore, the RAND analysis and analytic approaches had to remain quite flexible over this time. Here we summarize the important changes NAS made and how this affected NAS's theory of action.

Initially, its theory of action was quite simple and undeveloped: Designs would be developed, schools would adopt them in some unspecified manner, and this adoption would result in improved stu-

dent outcomes. NAS helped create designs through a Request for Proposals process from October 1991 to July 1992 in which more than 600 teams applied. It chose 11 design teams for initial development. From July 1992 to July 1993, these teams worked to further develop their theories of learning. A year later it reduced the number of teams from 11 to nine in part due to funding difficulties. It removed two teams that had a district as the team lead, indicating its theory of action did not include locally based and led teams in part because of NAS's ambition for national scale-up.

The demonstration phase, 1993–1995, took place in 147 schools chosen by design teams as partners. RAND analysis of progress at that point indicated that an important component of the design team intervention was the assistance teams provided to schools to enable the schools to adopt designs. This became known as “design-based assistance.” NAS's theory of change transformed to include this notion—that design teams did not just provide designs, but also had to provide design-based assistance for schools to successfully implement the designs. RAND analysis of progress in the demonstration schools also showed that school- and district-level factors had a strong relationship to implementation in schools. NAS then understood that to succeed, it would have to ensure a supportive environment for its designs. During this demonstration phase, NAS removed another two teams, both of which did not have national scale-up intentions and which appeared at the time to be closely associated with conditions specific to districts or states.

NAS outlined its scale-up strategy—a third iteration of its theory of action. Because of its experiences during the demonstration phase, NAS became more cognizant of the importance of gaining district-level support for teams and providing a supportive district structure, including resources, school autonomy, and professional development for the schools. Thus, it chose a concentration strategy—attempting to transform a few districts through the adoption of design-based assistance by a significant number of schools within those districts. Ten jurisdictions agreed to partner with NAS. These jurisdictions promised to get approximately 30 percent of their schools using designs within a three-year period and to provide a supportive environment for the schools' efforts. The 30 percent figure was admittedly somewhat arbitrary, but the theory of action was that if NAS could implement its designs in about one-third of a

district's schools, there would be a critical core of schools that could be sustained over time. In addition, NAS insisted the design teams become more self-sufficient at this point. Design teams would charge the districts fees for the design-based assistance. NAS would work with districts to help a supportive environment evolve. NAS promised districts that by using its designs, the schools would be able to show dramatic test score improvements within that time period.

The seven teams entering scale-up included:

- Audrey Cohen College (AC) (currently renamed Purpose-Centered Education);
- Authentic Teaching, Learning, and Assessment for All Students (ATLAS or AT);
- Co-NECT Schools (CON);
- Expeditionary Learning Outward Bound (EL);
- Modern Red Schoolhouse (MRSH);
- National Alliance for Restructuring Education (NARE) (currently renamed America's Choice Design Network); and
- Roots & Wings (RW).

NAS partnered with the following jurisdictions: Cincinnati, Ohio; Dade County, Florida; Kentucky; Maryland; Memphis, Tennessee; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; San Antonio, Texas; San Diego, California; and three districts in Washington state. About 185 schools partnered with designs in these jurisdictions, while as a whole, NAS design teams spread to over 550 schools by 1995.

By 1999, NAS design teams had over 1,000 partnering schools across the country. Since the scale-up phase, NAS has effectively and successfully been a proponent for the creation of a federal program to provide schools with funding to adopt designs. The Comprehensive School Reform Demonstration program was created in 1997 to provide schools with funding to implement designs similar to those created and developed by NAS, thus ensuring a market for NAS-like

designs. This has allowed NAS designs to spread to over 4,000 schools by 2001.

CONCEPTUAL FRAMEWORK

The overarching concept underlying NAS is the development of an intervention by external change agents who provide assistance during the implementation process in order to improve schools and student outcomes. In this case, the level of implementation achieved by the schools and teams working together is one important outcome. The other is the changes in student performance associated with adoption and implementation. These two concepts form the dependent variables in all the RAND analyses of NAS.

Past research on external change agents as a reform mechanism in K–12 education has shown that as these externally developed interventions are implemented, they tend to go through significant changes over time as they adapt to local conditions and contexts or in the process of scaling up. Often they develop implementation assistance strategies to assist schools in understanding and implementing the intervention. Despite this, implementation tends to be variable across sites, and the outcomes—in terms of the desired change—also vary considerably.

The conceptual framework underpinning the RAND studies of NAS draws from previous research on implementation of school reforms and educational change. A critical assumption underlying the designs is that coherent, focused, and sustained implementation of key design components (including professional development, curriculum and instructional materials, content and performance standards, assessments, organization and governance, and parent and community involvement) will eventually change school and classroom learning environments and thereby students' academic outcomes. Implementation consists of the process of putting into practice the elements or set of activities defined by design teams as core components of their design.

However, throughout the history of educational reform efforts, a prominent theme has emerged that the process of planned educational change is much more complex than initially anticipated (Fullan, 2001; McLaughlin, 1990). This is largely because of the

number of players involved and the number of factors that need to be aligned to support such fundamental change.

The factors that could be expected to affect both implementation and outcomes include the following, some of which were not readily controlled by NAS or its design teams:

- The design itself and its ability to offer coherent, comprehensive, and consistent education programs as well as assistance offered by the design teams to schools to ensure implementation;
- The efficacy of the selection and matching process between designs and schools to ensure teacher “buy-in” to the design;
- The capacity of the specific schools for undertaking the reform including the schools’ other efforts at reform, educational leadership, and teaching capability;
- School-specific demographics, structure, and climate;
- District contexts including the existing infrastructure supports and incentives for design implementation and improved student performance; and
- Other factors such as state contexts of testing and accountability, community contexts, and NAS funding policies.

The relationship between the independent and dependent variables is shown in Figure S.1.

METHODS AND CAVEATS

At the request of NAS (the client), RAND undertook different analyses at different periods of time that together amounted to a program of analytic studies on NAS. This program of analyses is summarized in Table S.1. The table highlights the purpose of each study, the approach taken, the sample used, the data sources, the measures developed, the contribution to the overall assessment, and the publications produced. (More information on methodology is provided in each chapter and in the Appendix as well as the accompanying RAND reports.)

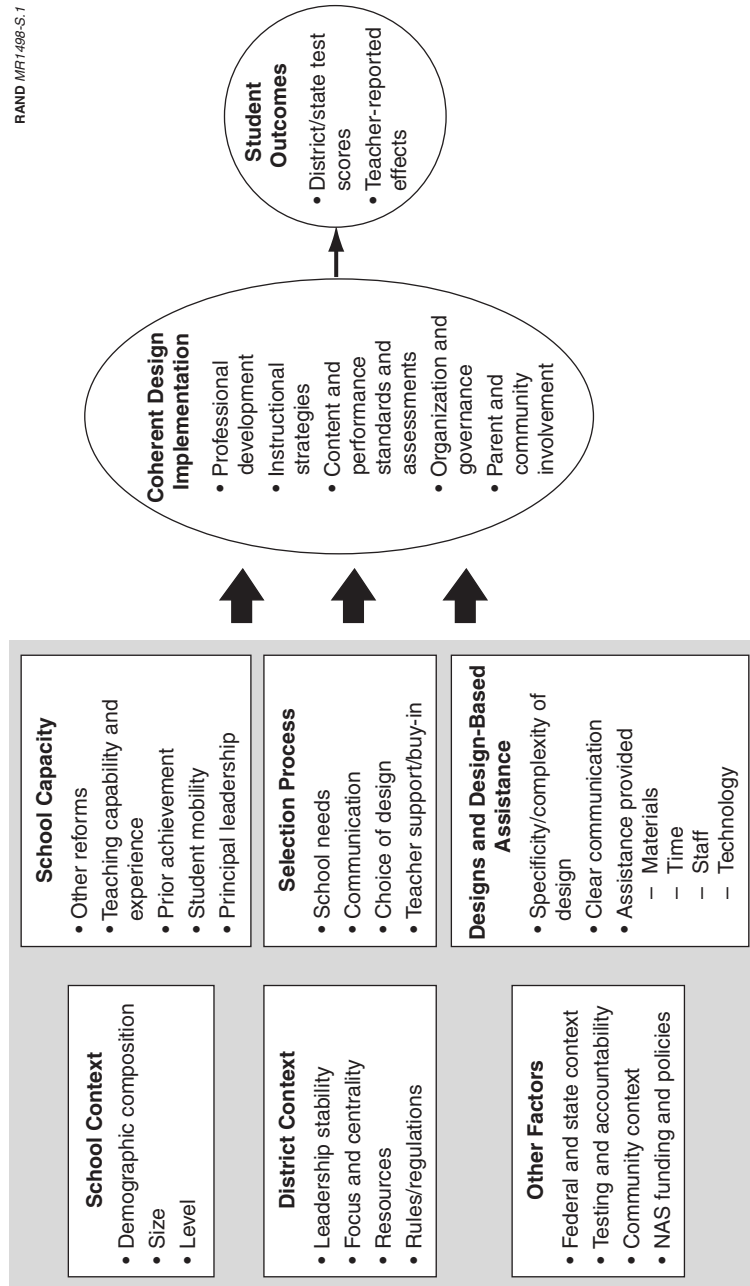


Figure S.1—A Conceptual Framework for Analyzing Implementation Progress and Performance in NAS Schools

These analyses can be thought of as “action-based research” during the developmental phase of an intervention. The RAND research on NAS was consistent with such an approach in that the program of studies:

- Allowed for a systemic view of the initiative from the teachers and students implementing the reforms to the district administrators trying to support them to design teams and NAS decisionmakers attempting to improve the reforms;
- Provided early information from front-line implementers on how to improve the effort; and
- Included a variety of methods, measures, and analytic techniques that were adapted to a changing environment to provide information that was relevant to NAS’s evolving theory of action.

The research and analysis was undertaken on behalf of a client involved in an experiment to engineer strong designs for schools. The use of the term experimental does not mean a controlled environment with random assignment of designs to districts or schools. It refers to NAS’s successive attempts to develop strong products called whole-school designs that would produce the outcomes NAS desired. This included successive efforts to learn what might have gone wrong in previous attempts, make adjustments, and try again. Therefore, the analyses reported here were adapted to the complex realities of the unfolding situation.

Each of the pieces of the RAND program of studies provides an important building block toward a full understanding of the NAS initiative. Together, they provide a cogent, consistent, and comprehensive examination of the NAS initiative and offer useful and timely information to decisionmakers considering or engaging in whole-school reform.

FINDINGS

We summarize the specific findings of the different analyses here. We group these findings into three broad areas. The first concerns designs and their development and refers in the conceptual frame-

Table S.1
Phases of NAS and RAND's Program of Studies

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Demonstration July 1993– July 1995	Compare & contrast designs; track progress & indicate issues	Longitudinal comparative case studies of design teams using qualitative analysis	9 teams, 32 schools, at least 2 for each team; National Alliance sites included schools at all grade levels within a district	Document review; interviews with design leads; survey of principals; nested interviews for each site from district to teachers two years in a row; site observations	No measures; Identify elements of design & levels of complexity of designs by coverage of elements & needed interactions to attain implementation	Identified notions of process vs. product teams; design specificity; design-based assistance; implementation issues; lack of district support	Bodilly et al., 1995; Bodilly, 1996; Mitchell, 1996

Table S.1—continued

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Scale-up July 1995– July 1997	Measure & analyze progress toward implementation	Longitudinal comparative case studies of design teams using qualitative analysis	7 teams 40 schools, at least 4 schools each, including elementary & secondary as possible; random selection	Document review; interviews with design leads; survey of principals; nested interviews for each site from district to teachers 2 years in a row; site observations; resources analysis	Implementation index based on progress toward design-specific elements	Showed contrast in progress by design & by district support; raised quality assurance issues; indicated types & quantity of resources required	Bodilly, 1998; Keltner, 1998; Bodilly & Berends, 1999

Table S.1—continued

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Scale-up March 1997– June 1999	Measure & analyze progress toward implementation & improved performance	Quantitative analysis of longitudinal sample of schools	7 teams; target sample of 184 schools in 7 districts actively implementing designs; complete longitudinal response from 71 schools from 1997–1999	Survey of all principals; survey of teachers; collection of quantitative outcome data	Two indices of implementation based on teachers' reports; a core index that included five common elements across all designs; a design team-specific index that included elements unique to specific designs; school level measures of performance based on district-wide mandated tests	Identified factors affecting implementation & performance including principal leadership, resources, teacher support for design teams, design team capacity & district policies	Berends, 1999, 2000; Berends & Kirby et al., 2001; Kirby, Berends, & Naftel, 2001

Table S.1—continued

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Scale-up April 1998– September 1999	Analyze factors leading to classroom implementation & effects on student achievement	Quantitative & qualitative analyses of NAS & non-NAS elementary classrooms within a high-poverty district	All 64 elementary schools in San Antonio district, 279 classrooms, 3,823 students; RAND sample of 23 schools; 861 students with wide array of data	Survey teachers; classroom observations; interview teachers, instructional guides, school & district staff; collect teacher logs & samples of student work; administer Stanford-9 open-ended reading test to 4th graders; district longitudinal data on student demographics & test scores, teacher & school characteristics	Student level achievement on RAND-administered test and state-mandated tests; classroom instruction practices consistent with designs	Identified complexity of reform progress & role of district in high-stakes environment & high-poverty, low-performing schools	Berends et al., 2002

Table S.1—continued

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Scale-up March 1999– June 1999	Analyze factors leading to schools with high implementation, but low performance outcomes	Case studies of schools indicated by teams as highly implementing	7 matched pairs of elementary schools; matched by district, demographics & purported levels of implementation	Level of implementation from designs; nested inter-views for each site from district to teachers	Consistency between observed level of implementation and that reported by design teams	Inability to obtain target sample precluded strong analysis; pointed to concerns over design team knowledge of site conditions & implementation issues; indicated multiple factors affecting performance other than design	Chun, Gill, & Heilbrunn, 2001

Table S.1—continued

Phase/Time	Purpose	Approach	Sample	Data Sources	Measures	Contribution	RAND Publications
Ongoing research October 1992– October 1998	Track & analyze changes to designs & design teams & NAS	Longitudinal case study analysis of designs starting with 11 & declining to 7; longitudinal analysis of NAS development	NAS and designs over period	Document review; interviews with design teams; review by teams & NAS; interviews with NAS staff; observations at NAS conferences, meetings, and briefings	Analyzed reasons for changes including: planned development, gradual adaptation to meet needs of sites, adaptation to conflicting policies	Showed significant changes in teams over time; points to importance of district, teacher, & student factors as causative in evolution of designs	Glennan, 1998; Bodilly, 2001

work chart (Figure S.1) to the box labeled “Designs and Design-Based Assistance.” The second concerns the ability of NAS and the design teams to implement the designs in schools. It first covers the level of implementation observed (i.e., the measure of the dependent variable), and then discusses the factors related to implementation (i.e., independent variables). Finally, we address student outcomes as an additional set of dependent variables as well as the related factors.

Overview of Changes to Designs Over Time

These findings are based on a continual tracking of the designs and teams throughout the analytic effort (Bodilly, 2001). Our purpose was to understand what intervention the schools attempted to adopt. It is based on a longitudinal case study analysis of each team. The analysis is based on a document review, interviews with teams, and reviews by NAS and the teams of the findings. The original design proposals were used as a baseline and changes to teams and designs were assessed against those original proposals. Interviews helped us understand why the changes were made.

Consistent with the literature on external change agents, the designs adapted over time. While some of the design development was beneficial for enabling schools to improve, other developments appeared less likely to help schools. For example, the growth in the assistance packages, the further development of curricular units, and the development of protocols for school choice of design all appeared on the surface to be positive adaptations. The development of basic skills curriculum could also be considered positive when well-integrated with the principles of the design and not simply a quick add-on to meet district demands.

Other changes, while understandable, remained more problematic. Adaptation to district and school policies led some designs to accept unaligned and incoherent mixes of standards, assessments, curriculum, instruction, and professional development. This also allowed for considerable local adaptation to the point where one might question the unifying nature of a design.

These changes to some designs over the demonstration and scale-up phases meant that a major component of NAS’s theory of action—a coherent, unifying design—was often missing or was constantly in

the process of being revised. It cannot be emphasized enough that during the entire time of the RAND studies, designs were still in a state of development. Consistent with the literature, implementation assistance—what NAS termed “design-based assistance”—became an important part of the intervention as time went on.

Overview of Implementation Findings

RAND studied the implementation of designs in both the demonstration and scale-up phases. The purpose was to measure the level of implementation in schools and to determine the conditions under which implementation prospered (Bodilly, 1998; Keltner, 1998; Bodilly and Berends, 1999; Berends, 2000; Berends and Kirby et al., 2001; Kirby, Berends, and Naftel, 2001; Berends et al., 2002; see Table S.1 for a brief descriptions of these studies).

The demonstration schools showed some promise, but RAND identified many barriers to implementation (Bodilly et al., 1995; Bodilly, 1996; Mitchell, 1996). During scale-up, NAS and the design teams partnered with schools and districts that were characterized by a host of problems related to poverty, achievement, and school and district climate characteristics (Berends, 1999). Achieving high levels of implementation within these schools across all teachers proved challenging. The case study analyses found that two years into implementation approximately half of the sample sites were implementing at a level consistent with NAS and design team expectations. The other half were below this level. All sites reported many barriers to further implementation.

The longitudinal sample of teachers supported the findings of the case studies showing lack of strong implementation and lack of increasing progress toward implementation:

- For the entire sample of schools we surveyed, implementation in the scale-up schools increased modestly from 1997 to 1999. The between-school variance in the level of implementation decreased somewhat over time, but the within-school variance increased. There was much greater variance in implementation *within* schools than between schools, suggesting that designs had failed to become “schoolwide.”

- There were large differences in implementation by jurisdiction, by design, and across schools.
- For schools newly adopting designs, implementation increased and deepened over the first four years after schools adopted designs, although at decreasing rates each year.

As expected, many factors influenced the level of implementation and the different analyses identified similar conditions leading to higher levels of implementation. The following summarized the key factors related to implementation as portrayed in our conceptual framework (Figure S.1) across all the RAND studies summarized in Table S.1.

Designs and Design-Based Assistance. Implementation varied by design. In the longitudinal survey sample, schools with CON, NARE, and RW designs clearly reported higher implementation. Reported high levels of implementation in all the studies were related to clear communication and strong assistance by design teams. These in turn brought about stronger teacher support for the designs in their schools. We remind the reader that in many cases designs were changing significantly over time and were still in the process of development during this entire period. Teachers reported poor communication of the design to the school staff at many sites because of the changing nature of the designs. Finally, in our last case study analyses of a small sample of schools we found that several designs did not have strong information about their schools nor had they adequately measured implementation for their schools. Taken together the evidence shows that several designs need to make significant improvements to ensure implementation of their designs in schools.

Selection Process. Our case study (Bodilly, 1998) revealed that those schools that felt they had a well-informed choice process reported higher levels of implementation than those that reported being forced to accept a design or not understanding the nature of the design. In the longitudinal survey sample, teacher support and buy-in were related to higher levels of implementation. Our case study work also revealed that principals played an important role in ensuring a sound selection process and the buy-in of teachers.

School Capacity. School capacity was important in supporting implementation. Principal leadership was an important contributor to implementation across all the studies. The survey analyses indicated teacher views of principal leadership was the most important indicator of the implementation level achieved. The teacher survey also indicated that teacher perceptions of students and their readiness to learn were all significantly related to level of implementation. The different case studies indicated that many other reforms were taking place in the schools while the NAS reform was under way. In many instances this caused teacher overload, and reduced the capacity of teachers to implement the designs.

School Context. The teacher surveys revealed that implementation was higher in high-poverty schools as well as schools serving high numbers of minority students. However, in schools that served significant numbers of *both* poor and minority students, implementation levels were significantly lower. Across our implementation studies, we found implementation was higher in elementary schools than in secondary schools and in smaller schools versus larger ones.

District Context. There were large differences in implementation between jurisdictions across all our studies. In general, implementation was higher in those districts that were more supportive of the NAS designs and characterized as having stable district leadership that backed the reform and made the reform central to its improvement efforts; a lack of crises such as budget crises or redistricting; a coherent program of reform; resources dedicated to the effort; significant school-level autonomy; and a trusting relationship between school, district and union staff. The several case study analyses highlighted significant barriers to implementation embedded in district and union policies.

The San Antonio classroom study gave particular insight into how district-level behaviors affected schools and their willingness to undertake and implement designs (Berends et al., 2002). Teachers' views of consistent and effective district leadership proved to be positively associated with teacher efforts toward implementation. Adoption of multiple reforms easily overwhelmed teachers and their efforts at this particular reform. Most importantly, the high-stakes state testing regime, which inadvertently encouraged a specific focus on basic skills, resulted in district adoption of specific curricular pro-

grams in addition to the designs. In fact, these curricular programs conflicted with the designs and resulted in lower levels of implementation.

Exit interviews with principals of schools that had dropped the NAS designs and related RAND research highlighted the importance of adequate resources in funding. Lack of funding was the single most important reason cited by most of the schools in the decision to drop a design. The costs of implementing the designs were far greater than the costs paid in fees to design teams. In fact, these fees made up less than one-third the cost to schools of implementing designs. Significant costs were borne by teachers in terms of teacher time and effort in implementing these designs.

Overview of Student Outcome Findings

We tracked school outcomes for those schools included in the scale-up phase, see Table S.1. Our analysis of performance trends across the set of schools three years into scale-up focused on whether NAS schools made gains—any gains—in test scores relative to their respective jurisdictions. While these school-level measures allowed us to compare performance in NAS schools with that of the district as a whole, they are subject to important limitations. For example, these aggregated measures may fail to capture changes in the tails of the distribution, or miss some significant achievement effects that may be captured if student-level data were available and comparable across jurisdictions. We found that:

- Among the four jurisdictions with 10 or more implementing NAS schools, Memphis and Kentucky schools appeared to be the most successful in terms of improvement in mathematics, while Cincinnati and Washington state did better in reading.
- In total, of the 163 schools for which we had data allowing comparisons in performance relative to the district or state, 81 schools (50 percent) made gains relative to the district in mathematics, and 76 schools (47 percent) made gains in reading.

Because of the wide variation in implementation and environments that occurred within schools and among jurisdictions, it may have been too early to expect robust performance results across the NAS

sites. Better and longer-term performance data are needed in order to make conclusive judgments about designs and their effects on school performance.

However, our implementation analysis showed that the overall level of implementation increased modestly over time, and there was continuing within-school variation in implementation. If the NAS approach to school improvement is correct, then weak implementation will lead to weak impacts on student performance. Our findings suggest that we cannot expect stronger performance results unless implementation significantly deepens.

The detailed classroom study of San Antonio allowed us to examine whether variation in instructional conditions was related to student achievement, controlling for other student, teacher, classroom, and school characteristics. We found that:

- Strong principal leadership as reported by teachers had significant positive effects on students' state test scores in reading and mathematics.
- Instructional conditions promoted by reforms such as NAS—including teacher-reported collaboration, quality of professional development, and reform-like instructional practices—were not related to student achievement net of other student and classroom conditions.
- In general, early implementation of NAS designs in a high-poverty district within a high-stakes accountability system did not result in significant effects on student achievement.

Other analyses of schoolwide reforms have found similar results.

CONCLUSIONS

In the end the RAND analysis illuminated three important areas. First, it provided evidence that an external agent such as NAS could deliberately create and promote design teams. Second, it indicated that some of the theory of action behind the NAS efforts was underdeveloped, and our various analyses pointed to important conditions for implementation and improved student outcomes. Third, it

pointed to important lessons about how to carry out future efforts at reform. We expand briefly here.

The Contribution of New American Schools

NAS did accomplish 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. These included:

- NAS funding and leadership led to the deliberate development of several functioning design teams.
- NAS showed that initially dependent external change agents could be moved toward self-sufficiency over time.
- NAS explicitly sought scale-up of the reform initiative.
- NAS actions as a change agent have significantly influenced policy in its areas of interest.
- NAS explicitly made analysis and good consumer education a part of its efforts.

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. NAS itself evolved from a venture capitalist organization to a provider of district- and design-based services.

NAS's Theory of Action

RAND findings provide mixed evidence to support NAS's theories of change.

The initial hypothesis, that by adopting a whole-school design a school could improve its performance, was largely unproven. We found specific positive examples of school implementation and improvement under certain conditions; however, negative examples were found under more common conditions. Our general findings showed difficulties in implementation and lack of strong improvements in school performance in a significant percent of the schools in our samples.

The hypothesis that designs alone were not helpful to schools and that schools needed assistance in implementation was proven correct. Teachers and school administrators clearly reported higher levels of implementation associated with strong assistance from design teams. 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. Without a supportive environment the designs did not flourish. Barriers to implementation reported by school staff focused on unsupportive district practices. They also pointed to the challenges inherent in adopting multiple reforms, high-stakes testing regimes and designs simultaneously.

In general then, we conclude that NAS's initial theory of action was largely underdeveloped and unspecified in terms of ensuring a successful scale-up experience. 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.

Implications for Future Efforts

Based on our experience with NAS, we offer the following implications for future attempts at scaling up of school reforms.

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 im-

plemented nationally. 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. 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 during 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. This implies the design concept and NAS’s initiative did not focus on important dimensions of school improvement when attempting to increase schools’ effectiveness.

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. 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.

A key component of successful implementation is consistent, clear, and frequent communication and assistance between design developers and schools, particularly teachers. 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. In part, this is a resource issue—the push to scale up can leave developers with few resources for evaluation and cycles of self-improvement. 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 not see evaluation as either important or even necessary. But, unless systems for tracking progress in schools and understanding school-level concerns are created and used for improving the external intervention, 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. In the developmental phases of an intervention, however, 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 the reforms' goals. The high-stakes testing regimes currently in vogue and the overwhelming emphasis given to test scores on state- or district-mandated tests as *the* measure of improvement do not bode well for many innovative reform efforts.