Chapter Four

IMPLEMENTATION OF NAS DESIGNS DURING THE SCALE-UP PHASE

Sheila Nataraj Kirby, Mark Berends, Scott Naftel, Christopher McKelvey, Susan Bodilly, JoAn Chun, Brian Gill, Jodi Heilbrunn

This chapter examines implementation of designs during the scale-up phase. First, we provide an overview of the schools in which scale-up occurred in order to set the context for understanding implementation and performance. The second section describes the research questions and provides a brief summary of the methods used by the various studies of implementation during the scale-up phase. These methods are described in considerable detail in the Appendix. This section also lists some caveats on the findings of the longitudinal analysis. The third section presents our findings on the levels of implementation achieved by the scale-up schools. The final section describes some of the important factors that affected implementation and some of the reasons why some schools eventually dropped designs.

AN OVERVIEW OF NAS SCHOOLS

Understanding the progress of NAS sites, particularly in terms of implementation and their performance on achievement tests (discussed in the next chapter), requires an understanding of where the schools were before implementing a design. Most of the schools receiving design team assistance could be considered socially and academically disadvantaged in terms of poverty, racial-ethnic composition, climate, and student test scores. The NAS sites in our sample were below "average" when comparing a number of school characteristics with national norms (Berends, 1999). For example, as Table 4.1 shows, the NAS schools in Cincinnati, Dade, Memphis, Pittsburgh, and San Antonio were serving mostly poor student popu-

lations—over two-thirds of the students were eligible for free/reduced-price lunch at these NAS sites. Philadelphia used a more stringent measure of poverty (percentage of students receiving benefits under the Aid to Families with Dependent Children program); hence the somewhat lower poverty rate in these schools. Design teams in Kentucky and Washington state were assisting schools that were more affluent than the national average. If these latter schools were excluded from the sample, the school poverty composition of the NAS sample increases to 68 percent.

The NAS design teams in Cincinnati, Dade, Memphis, and San Antonio were assisting schools that had a vast majority of minority students. By contrast, the NAS schools in the states of Kentucky and Washington are mostly non-Hispanic white schools. If the Kentucky and Washington schools were removed from the sample, over 80 percent of the students in NAS schools would be minority.

When examining implementation and performance of the NAS schools, it is important to consider the particular challenges that NAS design teams face when implementing their design. For the 104 schools that constitute the sample for our implementation analyses (see the Appendix for the derivation of the sample), Table 4.2 shows selected school characteristics by design team. MRSH and RW tend

Table 4.1

Comparison of School Composition: NAS Schools Versus Jurisdiction Schools, 1994–1995

	Percent Free/ Reduced-Price Lunch		Percent Minority	
	NAS Average	Jurisdiction Average	NAS Average	Jurisdiction Average
Cincinnati	74.9	58.0[a]	71.3	69.0
Dade	83.6	59.3	95.3	87.0
Kentucky	50.4	40.3	23.2	11.1
Memphis	80.2	66.0	89.6	86.0
Philadelphia	68.1	42.0[a]	56.4	80.0
San Antonio	99.2	91.1	95.0	95.0
Washington state	8.5	10.3[a]	13.2	16.3

[a]Data obtained from Common Core of Data for students on free lunch only.

to be in the poorest schools, while AC and RW tend to be in the schools with the highest percentage of minority students.

High student mobility is likely to have an adverse effect on implementation as well as school performance. As is evident from Table 4.2, highly mobile student populations characterize many of the schools that design teams are assisting. For example, nearly one in five students in RW and CON schools is likely to move during the academic school year.[1]

The distribution of the 104 schools across levels reveals that 64 percent are elementary schools, 14 percent are middle schools, and 14 percent are high schools. Eight percent are mixed levels.

In terms of school climate, NAS principals reported greater problems with absenteeism and school readiness when compared with the nation's principals (for details see Berends, 1999). School readiness included principal reports about problems such as students coming to school unprepared to learn, poor nutrition, poor student health, student apathy, and lack of academic challenge.

Table 4.2

Selected School Characteristics, by Design Team

	Free/Reduced-Price Lunch	Minority	Mobility	English Language Learners	Number
	Percent				
AC	79.6	95.8	7.8	2.8	5
AT	47.2	50.3	14.9	4.4	17
CON	71.9	80.6	19.3	1.2	12
EL	82.9	80.8	17.8	4.9	16
MRSH	88.1	84.3	13.4	3.9	7
NARE	40.3	19.5	10.3	2.0	32
RW	88.2	88.9	20.0	0.1	15
NAS Average	63.9	59.0	14.5	2.6	104

[1]Mobility rates are based on the following question in the principal survey: "On average, what percentage of your total student body enrolled at the beginning of the school year are still enrolled at the end of the school year?" Percentages in Table 3.5 are calculated as 100 minus this reported percentage.

In general, our data indicated that the majority of NAS sites were located in low-performing, urban school districts. Not surprisingly, within these districts and with few exceptions, the NAS design teams began assisting schools that were scoring at or below the district average on the district- or state-mandated tests.

STUDIES OF IMPLEMENTATION: RESEARCH QUESTIONS, METHODOLOGY, AND LIMITATIONS

In 1995, RAND began an evaluation of the scale-up of NAS designs to many schools. The longitudinal evaluation of the scale-up phase covers years 1995 to 2000 and addresses five major questions:

- What was the level of implementation of NAS designs across this set of early-implementing NAS schools two to five years after scale-up? Has implementation increased over time? Does implementation differ by jurisdiction and design team?
- Has implementation deepened over time across schools, as measured by the change in the within-school and between-school variance of reported implementation levels between schools?
- What are the factors—in terms of teacher, school, design team, and district characteristics—that help explain the variation in implementation across schools and jurisdictions?
- Among schools that dropped the NAS designs and for which we have data, what factors contributed to this decision?
- Does the adoption of NAS designs result in any changes to student and school outcomes?

Our findings with respect to questions 1–4 are addressed in this chapter. Question 5 is the focus of Chapter Six.

As we showed in Chapter One, RAND's program of implementation studies has included:

- 1996–1997: Case studies in 40 schools two years into scale-up to analyze implementation and the role that districts play in impeding or enabling comprehensive school reform (Bodilly, 1998);

- 1995–1999: A longitudinal analysis of between 70 and 100 NAS schools that began implementing early on in the scale-up phase, for which data on implementation and performance were gathered from principals, teachers, and districts (Berends and Kirby et al., 2001; Kirby, Berends, and Naftel, 2001); data collected from these schools cover the period two to four years after scale-up;[2] and
- 1999: A case study analysis of what factors contributed to performance differences in high-implementing NAS sites five years after scale-up, using a matched set of schools (matched on the basis of design, district, grade span, years of implementation, and implementation level, as measured by our surveys but validated by the design teams). One school was high performing and the other was not.

The methodology for each of these studies is discussed in some detail in the Appendix.

Limitations of the Study

It is important to understand the limitations of our sample and findings drawn from analyses of this sample of schools. For many of the design teams, these were the first schools to which they had pro-

[2]In addition, because the longitudinal sample focused on early-implementing schools, RAND collected data from a freshened sample of schools that began implementing NAS designs after 1995–96. However only four jurisdictions—Cincinnati, Memphis, San Antonio, and Washington state—agreed to participate in this data collection effort, and 46 schools in these jurisdictions responded to the principal and teacher surveys. Although we analyzed data from these schools, the analyses did not substantially change the results from those reported here. These results were not included in the set of reports on implementation, but provide some assurance of the robustness of our findings.

vided assistance with implementing their designs on a fee-for-service basis. In addition, at the beginning of scale-up in 1995, most of the design teams reported to RAND that their designs were still unfinished. As a result, the early years of implementation on which we report saw many changes in both the designs and the assistance provided as the teams and the schools gained experience.

The strategy that NAS developed for scale-up (NASDC, 1997) focused on a small number of jurisdictions that persuaded NAS that they possessed what NAS called "supportive operating environments" in which the designs could be implemented. In fact, for the most part, these districts did not possess such environments. They had limited understanding of whole-school reform and the sort of design-based assistance that NAS design teams were intending to provide. The districts, NAS, and the design teams collectively and individually invented procedures and policies for design teams and the assistance they provided as the implementation unfolded. For example, districts varied widely in the processes set up for matching schools and designs, the contracts set up with designs, the services to be acquired, and the ways they monitored implementation of the designs (Bodilly and Berends, 1999; Bodilly, 1998).

In short, the early years of scale-up continued to be a time of uncertainty. There was some chaos and a great deal to be learned on the part of NAS, designs, districts, and schools. Thus, this report documents experiences that may differ from those of schools beginning implementation today. NAS and the design teams might have matured due in large part to the lessons learned about the ways in which jurisdictions, design teams, and schools must work together (Bodilly and Berends, 1999; Bodilly, 1998).

While the fact that designs were evolving over time as they gained experience and adapted to local contexts makes a longitudinal evaluation difficult, we believe that the information obtained in following these schools still offers valuable lessons, particularly for CSRD schools adopting a variety of school-reform models in many differing environments. Thus, when interpreting the implementation findings, it is important to keep in mind these features of the population of schools we have studied.

IMPLEMENTATION LEVELS IN NAS SCHOOLS

Measuring Implementation in the Case Study Analysis

The study created an implementation scale based on common elements of the designs. These common elements were curriculum, instruction, assessments, student assignments, and professional development. We tracked progress on all these elements, as applicable. We rated progress in an element using a straightforward scale, as follows:

0 = Not Implementing. No evidence of the element.

1 = Planning. The school was planning to or preparing to implement.

2 = Piloting. The element was being partially implemented with only a small group of teachers or students involved.

3 = Implementing. The majority of teachers were implementing the element, and the element was more fully developed in accordance with descriptions by the team.

4 = Fulfilling. The element was evident across the school and was fully developed in accordance with the design teams' descriptions. Signs of institutionalization were evident.

Constructing a Core Implementation Index[3]

In our longitudinal analyses, we used a core implementation index to measure the average level of implementation in NAS schools. The core implementation index is a summative scale of teacher responses as to the degree to which the following described their school (on a

[3]In our earlier report, Berends and Kirby et al. (2001), we also developed a design team–specific implementation index that measures implementation of both shared and some unique aspects of the designs. The design team–specific index allowed us to measure implementation of each design on components that are unique to and emphasized by the design. The shortcoming of this index is that it is not directly comparable across designs because it varies both in terms of items and number of items included in the index. The details of this index are provided in the Appendix .

scale of 1–6, with 1 = does not describe my school, and 6 = clearly describes my school):[4]

- Parents and community members are involved in the educational program;
- Student assessments are explicitly linked to academic standards;
- Teachers develop and monitor student progress with personalized, individualized learning programs;
- Student grouping is fluid, multiage, or multiyear;
- Teachers are continual learners and team members through professional development, common planning, and collaboration; and
- Performance expectations are made explicit to students so that they can track their progress over time.

Teacher responses were averaged across a school to obtain the school mean level of implementation. We analyze this overall implementation measure for two reasons:

First, the core function of schools is teaching and learning. Therefore, we selected those teacher-reported implementation indicators that were related more directly to influencing what goes on in teachers' lives and inside classrooms. From an organizational perspective, classroom instruction is the core technology of school organizations and the primary mechanism through which learning occurs (Gamoran et al., 1995; Gamoran and Dreeben, 1986; Parsons, 1959). It is this core function of schools that the designs ultimately want to influence and it is this aspect of implementation that our overall implementation index aims to measure.

Second, we want to examine factors related to implementation, and this summary measure allows us to present our results in a parsimonious manner.

---

[4]The alpha reliability of this index was 0.81. The range of correlations for the individual items was 0.21 to 0.57.

However, measuring progress in implementation broadly across a wide set of schools in several partnering jurisdictions involved a number of challenges (Berends and Kirby et al., 2001; Kirby, Berends, and Naftel, 2001), including the uniqueness of the designs and the fact that the designs were still evolving. The Appendix contains a more detailed discussion of these issues.

In the analysis sample of NAS schools that we examine, small sample sizes for some design teams make traditional tests of statistical significance somewhat more difficult to apply. That is, with larger sample sizes we would have more power to detect differences and effects. Thus, we focus on what appear to be educationally substantive differences where appropriate.

Despite these challenges, evaluation remains an important component of any effort to change schools, and it is important to develop and refine sets of indicators that are informative not only for researchers, but for design teams, educators, and policymakers.

Thus, in order to address our questions about implementation stated above, we developed the core implementation index described earlier to broadly measure implementation of the *major*, shared components of the designs across the sites. The core implementation index is useful for understanding the progress of NAS schools during the scale-up phase.

Findings

In general, no matter which method we used—case study analysis or the core implementation index—we found that implementation levels were less than ideal in schools adopting the NAS designs.

Two Years After Scale-Up. RAND's early case studies found that schools varied considerably in the level of implementation achieved two years into the five-year scale-up effort. *Generally, about half of the 40 schools examined in the case study research were implementing at targeted levels (levels desired by teams, NAS, and districts) while the other half were below this level (Bodilly, 1998). The level of implementation varied by design team, district, and school characteristics.*

Four Years After Scale-Up. Schools responding to the longitudinal surveys showed similar findings with less-than-high levels of implementation, even after four years into the scale-up period. The following indicates the findings from the survey sample:

Implementation increased modestly from 1997 to 1999. The between-school variance decreased somewhat over time, and the within-school variance increased. In order to make it easier for the reader to gauge the magnitude of the changes over time, we calculated standardized z-scores based on the mean and standard deviation of the 1997 core implementation index. This allows us to represent changes using a common metric. The mean implementation index was 4.14 in 1997 with a standard deviation of 0.61. Thus, the standardized mean for 1997 is zero, with a standard deviation of 1.

Figure 4.1 shows the distribution of the core implementation index for all 71 schools in the longitudinal sample across the three years of data, using a standardized z-score based on the mean and standard deviation of the 1997 core implementation index.[5] The mean implementation index rose modestly by about 0.25 of a standard deviation in 1998, and by 0.29 of a standard deviation in 1999. The difference between 1997 and 1999 was statistically significant, using a paired t-test for means.[6]

The spread declined over time as well, as can be seen from the figure. Although not shown here, the variance in mean implementation among schools declined over time. The standard deviation declined from 0.61 in 1997 to 0.57 in 1998 and 0.52 in 1999. This decline was

---

[5]This graph and some of the others that follow are portrayed with box-and-whisker diagrams, which show the distribution of the particular indicator being examined. In a box-and-whisker diagram, the line in the box is at the median value—half the values fall above the line and half fall below. Each "box" captures the middle 50 percent of the distribution. The lines, called "whiskers," at each end of the box show the range of scores beyond the upper and lower quartiles. Outliers are indicated by the shaded circles. The box-and-whisker plot thus allows us to compare the centers (median or center of the box), spread (measured by the interquartile range or the height of the box), and tails of the different distributions.

[6]These are calculated as follows: The mean implementation index was 4.29 in 1998; thus the z-score for the 1998 mean is $(4.29 - 4.14)/0.610 \approx 0.25$. Similarly, the z-score for the 1999 mean is $(4.32 - 4.14)/0.610 \approx 0.29$.

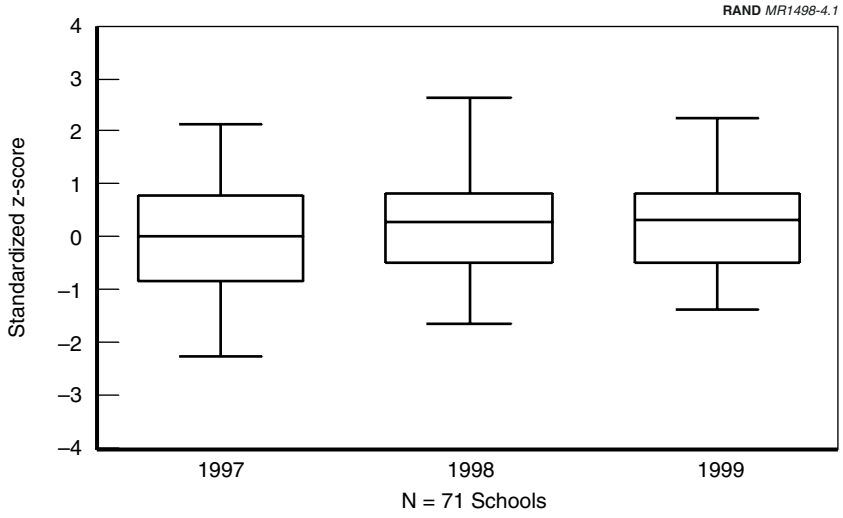


Figure 4.1—Standardized Z-Scores of the Overall Implementation Index (Based on 1997 Mean and Standard Deviation), 1997–1999

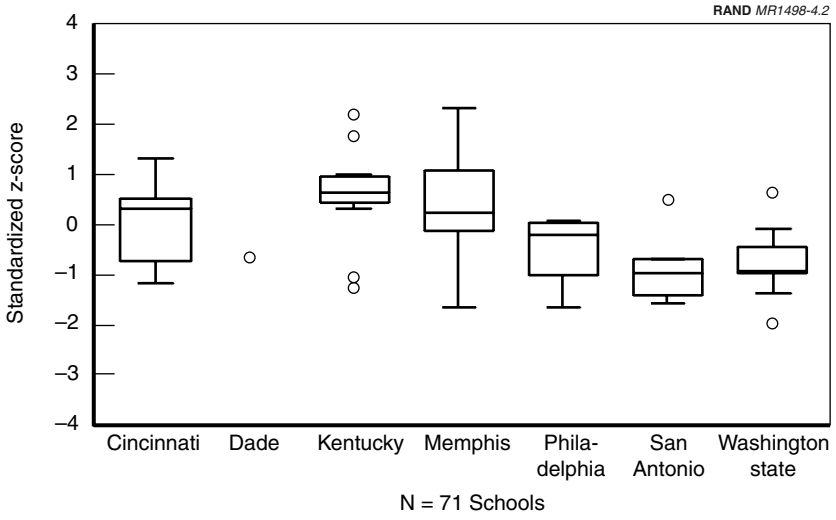
not statistically significant. However, the within-school variance increased over the same time period, suggesting that implementation did not become more "schoolwide" within a school.

In our multivariate analyses, we decomposed the variance in implementation into its variance components: within-school variance and between-school variance. The variance in implementation within schools was much larger than the variance between schools. In fact, only 18 percent of the total variance in reported teacher implementation was between schools; the remaining 82 percent was within schools. The between-school variance component declined from 27 percent in 1998 to 18 percent in 1999, with a corresponding increase in the within-school variance component. Such findings are not uncommon in analyses of school contextual effects on student and teacher outcomes (see Lee and Bryk, 1989; Gamoran, 1992; Berends and Kirby et al., 2001). However, because of such differences within schools, educators, design teams, and policymakers may need to think carefully about how to implement changes throughout the school.

Our multilevel models explained almost all of the between-school variance and about 31 percent of the within-school variance.

There were large differences in implementation by jurisdiction in 1999. We found large differences in the distribution of the core implementation index across the jurisdictions as well as design teams (Figure 4.2). In 1999, the mean implementation index was 4.32, with a standard deviation of 0.52. We calculated a standardized z-score for each jurisdiction and each design team, based on the 1999 mean and standard deviation for all schools.

Kentucky and Memphis ranked relatively high on this index with means that were 0.60 and 0.33 of a standard deviation higher than the mean while Washington state and San Antonio ranked the lowest, with means that were 0.77 and 0.87 of a standard deviation



NOTE: Sample sizes are Cincinnati (n = 10), Dade (n = 1), Kentucky (n = 13), Memphis (n = 24), Philadelphia (n = 6), San Antonio (n = 7), Washington state (n = 10).

Figure 4.2—Core Implementation Index (Standardized Z-Scores), by Jurisdiction, Spring 1999

lower than the mean.[7] Memphis schools also displayed the greatest spread in the data, as is evident from the long whiskers in the figure. Kentucky had a number of outliers, both high and low. Cincinnati also showed a great deal of spread, with schools having means that ranged from well below one standard deviation below the overall mean to well above one standard deviation above the overall mean. The differences between the highest and lowest jurisdictions were all statistically significant.[8]

There were large differences in implementation by design teams in 1999. Comparisons among design teams reveal that CON, RW, and NARE ranked comparatively high on the core implementation index while MRSH generally ranked the lowest, reflecting the ranking we found in 1998 (Figure 4.3). CON schools had a mean that was almost one standard deviation higher than the overall school mean while RW and NARE schools had means that were two-tenths and one-tenth of a standard deviation higher than the overall mean. MRSH schools had the lowest mean, over half a standard deviation below the overall mean. However, in terms of differences in means, none of these differences was statistically significant.

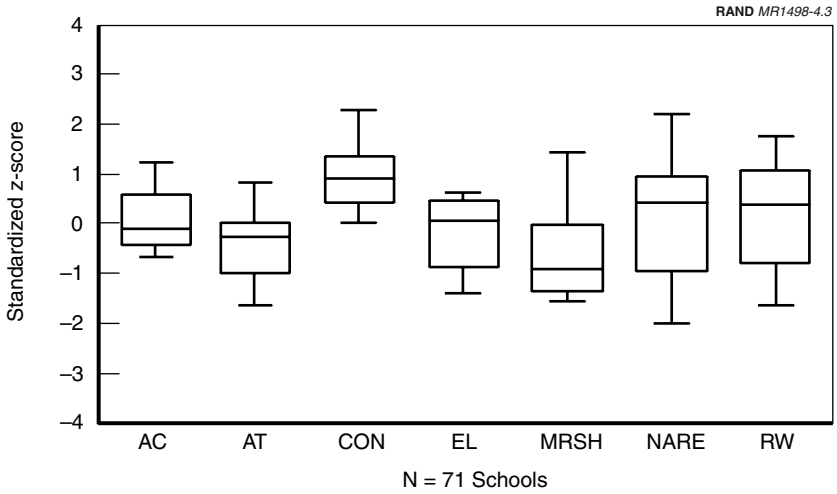
Implementation appeared to increase and deepen over the first four years after schools adopted designs, although at a decreasing rate. Figure 4.4 summarizes the actual relationship between years of implementation and the level of implementation for schools in our sample.[9] We see a sharp increase between the first and second years,

---

[7]The following is an example of how these effect sizes are calculated: The mean implementation index for Kentucky was 4.63; thus the z-score for Kentucky is $(4.63 - 4.32)/0.52 \approx 0.60$.

[8]Statistically significant here refers to the mean differences being significant at the 0.05 probability level or less. This is based on the multiple comparison test using the Bonferroni correction.

[9]In calculating the mean level of implementation for each group, we had more than one data point for some groups, based on the three years of data. For example, schools that had been implementing for one year in 1997 had been implementing for two years in 1998. We also had some schools that had been implementing for two years in 1997. In such cases, we used a weighted average of the mean level of implementation reported by these two groups of schools, where the weights were the number of schools in each group.



NOTE: Sample sizes are AC (n = 4), AT (n = 13), CON (n = 6), EL (n = 10), MRSH (n = 7), NARE (n = 19), RW (n = 12).

Figure 4.3—Core Implementation Index (Standardized Z-Scores), by Design Team, Spring 1999

modest increases from second through fourth years, and a sharp decrease in the fifth year. Schools with more than five years show higher levels of implementation, although the sample sizes are quite small.

Five Years After Scale-Up. *Even in schools selected by the* design teams *as high implementing, the level of implementation was quite low five years after scale-up.* In the course of gathering our sample, we learned that the various design teams, except RW, were not as knowledgeable about their schools as one might expect given their focus on design-based assistance. Information regarding schools' implementation levels was often outdated, overly optimistic, or simply missing. Although we had requested a sample of high-implementing schools, the levels of implementation across the

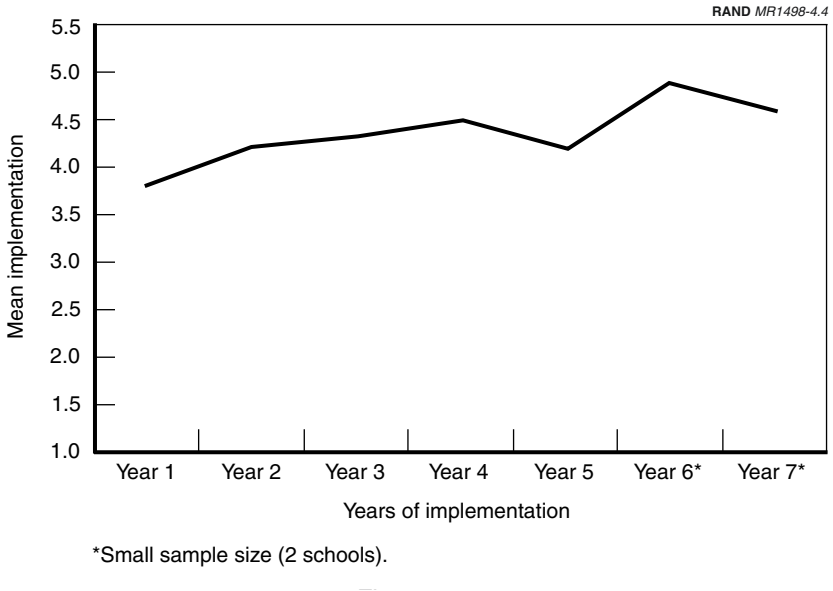


Figure 4.4—Mean Implementation, by Years of Implementation, Longitudinal Sample

schools we visited tended to be low. The low levels of implementation were not a consequence of the newness of designs to schools because most schools in our sample had adopted their designs at least three years ago.

FACTORS AFFECTING IMPLEMENTATION IN NAS SCHOOLS

Several factors emerge from both the case study and survey research as fostering high-quality and coherent implementation in the types of schools in the sample, perhaps the most important of which is principal leadership. The findings are grouped into broad categories, earlier identified in our conceptual framework.

Designs and Design-Based Assistance

Designs and the assistance they offered clearly affected the level of implementation achieved.

Type of Design. Bodilly (1998) found that teams that placed greater relative emphasis on core elements of schooling (curriculum, instruction, student assignment, student assessment, and professional development) rather than a more systemic approach tended to have higher implementation. Similarly, the longitudinal survey found that overall certain designs had markedly higher levels of implementation: CON, NARE, and RW, while others such as MRSH had markedly lower levels of implementation. In the multivariate model, controlling for other factors such as prior implementation and school characteristics, we do not find many differences among designs, with two exceptions: CON schools and AC schools made steady progress over this time period. AC schools, which in 1997 were at the low end of the implementation index, have made marked progress in implementation over the two years. This may be due to unobserved characteristics of the designs themselves that make them easier or harder to implement in schools already facing several challenges in terms of poverty, lack of resources, and the capacity to implement designs—a critical issue for future research to address.

Of 13 schools that had been implementing for three or more years, implementation levels were higher in the RW schools, which implemented only the reading component of the RW design. RW schools achieved higher implementation levels than the other design schools in our sample because RW provided schools with virtually all of the necessary curriculum and pedagogy, requiring less initiative from teachers. It also provided frequent, consistent, and reliable implementation checks. Finally, RW, as a reading program rather than a truly comprehensive school design, was far less ambitious an initiative to take on than the other designs.

Importance of Clear Communication. Our findings highlight the importance of clear communication to teachers in facilitating higher implementation (Bodilly, 1998; Berends and Kirby et al., 2001; Kirby, Berends, and Naftel, 2001). Clear communication had a large and statistically significant effect on the level of implementation. Of course, this variable and the teacher support variable were correlated.

Design-Based Assistance. Bodilly (1998) found that two important contributors to design implementation were having a stable team with the capacity to serve a growing number of schools and design

team support in the form of resources in encouraging high levels of implementation. The longitudinal study found that on average, schools that reported more resources for implementation (e.g., materials; professional development; time for planning, collaboration, and development; consultants to provide ongoing support; technology; and funding) had higher levels of implementation (Berends and Kirby et al., 2001).

Chun, Gill, and Heilbrunn (2001) reported that teachers they spoke with received training and some design-based assistance, but found that neither consistently met needs. Some teachers reported that the training provided over time was not sufficient. Moreover, the quality of training varied by design representative.

School Capacity

Strong Principal Leadership. Schools reporting having strong principal leaders had implementation levels over half a standard deviation above schools at the sample average. In addition, individual teachers' beliefs about principal leadership were important in explaining within-school variance in implementation. Our findings suggest that effective and supportive principal leaders are likely to both increase and deepen implementation in a school. For example, if most or all the teachers in a school view the principal as a strong leader, this is likely to reduce the variance within a school and help the design become more schoolwide. The importance of principal leadership for establishing effective schools has been emphasized by researchers for decades (Edmonds, 1979; Purkey and Smith, 1983; Rosenholtz, 1985), so it is not surprising that such leadership is critical for the implementation of NAS designs. While not surprising, the crucial role that principal leadership plays with respect to implementation should not be overlooked when adopting and implementing whole-school reforms.

Teacher Factors. Teachers' characteristics and their attitudes were also important determinants of level of implementation across the different analyses. For example, in the longitudinal survey, we found that teacher perceptions of students and their readiness to learn were all significantly related to teacher-reported levels of implementation. Teachers with a greater sense of efficacy—i.e., those who believed strongly that lack of basic skills was not a hindrance to their

students' academic success, or that students can learn with the resources available—also reported higher implementation than those who felt otherwise. We acknowledge that teacher efficacy may not be entirely exogenous in our model—it is certainly plausible that higher implementation of designs may have increased teachers' capacity to work with their students and hence their sense of efficacy. If so, the relationship is not causal but correlational. Even so, we believe our findings underscore the importance of enhancing teachers' abilities to work in diverse settings and providing them with the resources and supports they need.

School Context

School Composition. Taking into account other factors related to teachers, design teams, and districts, we found that poverty and minority composition of students were related to implementation, both in a positive direction. Teacher-reported implementation levels were higher in higher-poverty schools and among schools with high percentages of minority students.[10] It is interesting and promising to find that schools serving largely poor or minority students reported more success at whole-school reform. This may be largely a question of motivation or determination to succeed on the part of the teachers and principals in these schools. It also offers an indication of the ability of some designs to help change these challenging schools. However, in our longitudinal analysis, the positive, separate effects of poverty and minority composition were largely wiped out in schools that ranked high on *both* poverty and minority composition.[11] Unfortunately, the small sample size prevented us from decompos-

---

[10]Schools that ranked 10 percentage points above the sample mean on either of these variables reported levels of implementation that were one-tenth of a standard deviation higher than schools at the sample mean.

[11]Because poverty and minority composition are strongly correlated (the correlation coefficient is 0.76), we introduced an interaction term to see whether the combined effect of high-poverty and high-minority composition was different from the effects of these two variables separately. The estimated effect of the interaction term was equal to the sum of the coefficients on poverty and minority separately and in the opposite direction; schools that were 10 percentage points higher in terms of poverty and minority composition, relative to the sample mean, reported implementation levels two-tenths of a standard deviation lower than schools at the sample mean. On net, the combined effect of poverty and minority composition (the interaction term) washes out the separate positive effects of these variables.

ing these results further, but it is clear that more work remains to be done to understand how these characteristics of schools affect implementation.

School Level. Bodilly (1998) also found that at the school level, implementation was higher in schools that were elementary schools; were well informed about the design they selected and allowed free choice of design; and were relatively free of strife and had stable leadership.[12]

Our discussion has focused thus far on the net influence of each factor. However, it is important to emphasize that schools often face a multiplicity of challenges, and the interaction among these factors can set these schools back considerably in their attempts to implement school designs. For example, Bodilly (1998) found that schools that were beset with a combination of two or more negative factors, such as internal tensions, leadership turnover, forced adoptions of designs, or poor understanding of designs, ranked very low on implementation. Thus, schools need stable leadership and capacity and commitment on the part of the teachers to make the designs work.

Selection Process

Teacher Support for the Model. This variable was important in explaining both within-school and between-school variance in implementation in the survey. Supportive teachers implemented at a higher level within a school; the greater the degree of overall school-

[12]Our earlier work (Berends and Kirby et al., 2001) also showed that some school demographics were related to implementation, notably size, school level, and student mobility, although the effect differed across the models. In the models where we did not control specifically for jurisdiction, characteristics of the schools appeared to be more important. Large schools had significantly lower levels of implementation (about one-fifth of a standard deviation lower). Secondary schools also reported lower levels of implementation, although the effect was significant in only one of the models. Student mobility had a negative impact on implementation, as one would expect. Schools with higher student mobility reported levels of implementation that were about one-tenth of a standard deviation lower that those with lower student mobility. In the later work (Kirby, Berends, and Naftel, 2001), these school demographics did not appear to be significant. One possible reason might be that, as time goes on, these school demographics do not play as important a role as leadership, teacher sense of efficacy and support for the model, and support from the design team and district.

level support, the higher the implementation. This highlights the importance of getting teachers behind the adopted model; supportive teachers tend to reinforce and enhance implementation, not merely at the individual teacher level but at the school level as well. Whether teachers voted to adopt the model was largely subsumed in the teacher support variable, and as such, did not have a separate effect on implementation. Similar results were found by Chun, Gill, and Heilbrunn (2001). Due to limited information and uneven design-based assistance, not all teachers ever fully understood or accepted their respective designs.

District Context

The district context also proved to be important in impeding or ensuring implementation. Our findings highlight the importance of stable district leadership, provision of adequate resources, and supportive rules and regulations.

Districts played several important roles in fostering/hindering implementation, including: initial matching and selection; encouraging support by the design team; and creating a supportive environment with political leadership, regulatory policies, and consistent funding stream (Berends, Bodilly, and Kirby, forthcoming).

At the district level, Bodilly (1998) identified several district and institutional factors that contributed to implementation. These were leadership backing and stability at the district level; centrality of the NAS initiative to the district's agenda; lack of crisis situations; history of trust and cooperation; availability of resources for transformation; school-level authority and/or autonomy; union support; district accountability; and assessment systems that were compatible with those of the designs.

The longitudinal survey (Berends and Kirby et al., 2001) showed that the level of implementation varied significantly across districts. In that analysis, we found that Memphis and Kentucky ranked high on these indicators of support and ranked high in implementation, while others, such as San Antonio and Washington state, lagged far behind.

In the case studies of 13 schools that had been implementing for three years or more, we found districts were supportive of their schools' efforts in the sense that they provided the necessary funds to implement their respective NAS designs. However:

- Most districts undermined their support to schools and teachers (perhaps unintentionally) by also requiring them to incorporate districtwide initiatives that in some cases conflicted with design approaches to curriculum and instruction;
- In addition, schools' capacity to carry out the reform was limited because districts had not granted schools the autonomy required by the designs;
- Following the lead of their districts more than the elements of their respective designs, teachers found themselves struggling to juggle multiple responsibilities and initiatives, resulting in even less time to learn about design features and engage in design-related activities.

FACTORS THAT CONTRIBUTED TO THE DECISION TO DROP THE DESIGN

As part of the longitudinal analysis, we conducted additional exit interviews with principals who reported dropping the NAS designs out of the original sample of 155 schools in either 1998 or 1999. We asked principals about the factors that contributed to the decision to drop the NAS design and what advice principals would give to schools on the verge of implementing a whole-school design. Thirty principals responded to the interviews,[13] and their responses offer some valuable insights. These responses also point to the importance of funding, supportive leadership at the district and state levels, and assistance from design teams in ensuring sustained implementation.

[13]It is difficult to calculate an attrition rate (i.e., schools that dropped the design as a percentage of the total sample) for the sample as a whole. Some schools that did not respond may well have dropped the design. Out of the 184 schools at the beginning of scale-up and excluding the 12 Pittsburgh schools that were later dropped from the study, at least 41 schools out of 172 had dropped the design, giving us a lower-bound attrition rate of approximately 24 percent.

Funding to pay for the design and for professional development for teachers was the primary reason for dropping the design, while lack of support from the district and the state ranked second. Schools were also unhappy about the amount of effort required on the part of teachers to implement the designs, and the materials or training provided by the design teams. Nine principals cited "other reasons," but on a closer reading, these reasons appeared to relate in some way to lack of funding, lack of support at the state level, and dissatisfaction with the assistance provided by the design teams.

About ten schools reported that they were planning to replace the design with some other reform effort. Seven schools reported that this new program was more curriculum-centered than the design, and four reported that this was another whole-school reform. Interestingly, although the schools were dropping the design, a little under half reported that they planned to continue some elements or aspects of the designs.

Principals were asked what advice they would give to schools that were considering adopting whole-school designs. Some of their comments are noteworthy:

> Make sure you know everything up front—costs, training, and ask for five-year commitments. Make arrangements for time.

> First consider what type of kids you are serving. . . . These programs do not fit every building. Make sure you have enough money for training. If you don't have the money, there's no use jumping into it.

> Do it systematically, be careful with selection. Make sure the selections are successful in a variety of different settings and have data to prove it. Make sure you have faculty buy-in.

> Make sure all teachers and stakeholders understand the need and design of the change. . . . After all, faculty is going to be doing it.

> Make sure you have funding and support from state and district.

> Research the design thoroughly. Visit schools that have implemented the program. Call the state and see how long they project the program's continuation.
>
> Be patient in seeing significant change. With staff, people accept change differently. Change takes time and new learning.

This advice resonates with findings discussed earlier.

SUMMARY AND POLICY IMPLICATIONS

NAS and the design teams partnered with schools and districts that were characterized by a host of problems related to poverty, achievement, and climate characteristics. To scale-up the designs or replicate implementation in these sites proved difficult.

Level of Implementation

Four years after scale-up, schools reported relatively modest levels of implementation, although average level of implementation across all NAS schools did increase from 1997–1999 (two to four years after scale-up). Achieving high levels of implementation, especially within a school by all teachers, proved challenging. Even four years after scale-up, there was considerable variance in reported implementation within a school. There were large differences in implementation by jurisdiction, by design, and across schools.

Factors Affecting Implementation

It is clear that several factors need to be aligned for designs to be well implemented in schools. Without strong principal leadership, without teachers who support the designs and have a strong sense of teacher efficacy, without district leadership and support, and without clear communication and provision of materials and staff support on the part of design teams, implementation is likely to lag far behind. These are sobering and important lessons for any efforts at school reform. They underscore the basic inequality among schools in terms of capacity to undertake reform and point to the need for development of leadership and staff capacity as the precursor to reform, not

necessarily the result of it. In some instances, schools and districts may need to adopt a two-tiered approach to implementation—building up school capacity, particularly the skills and readiness of both the staff and students to implement change, and then attempting to implement the whole design throughout the school.