This chapter addresses whether schools implemented critical components of the NAS designs. The earlier report (Berends, Kirby, et al., 2001) reported on the status of implementation of the NAS designs as of spring 1998; this report provides an update on the status of implementation a year later. As we showed earlier, by the spring of 1999, all of the schools in our longitudinal sample had been implementing for three or more years.

RESEARCH QUESTIONS

Our analyses focus on two sets of research questions:

1. **What is the mean level of implementation of NAS designs across this set of early implementing NAS schools three years after scale-up?** Has implementation increased over time? Does implementation differ by jurisdiction and design team? We should expect an increase in the mean level of implementation over time as implementation takes hold, but the rate of change in the level of implementation is likely to vary depending on whether schools are in the relatively early phase of implementation or are more experienced with the designs. For example, if—as seems likely—implementation follows a polynomial function, increasing sharply over the first few years and then leveling off, we would expect sharp increases in mean implementation in the first few years after adoption, as schools moved to adopt and deploy key components of the designs. After a few years, however, there should be a tailing off of the increases in mean implementation
levels, as designs become more schoolwide and more an integral part of the daily work life of principals, teachers, and students.

2. Has implementation deepened over time both within and between schools, as measured by the change in the variance of reported implementation levels within and between schools? As implementation deepens, we should expect a decrease in the variance of the implementation index both between and within schools, as designs become more schoolwide, and there is greater consensus, clarity, and coherence in what teachers within a school are doing. This latter effect will be tempered by jurisdictional effects because of the varying degrees of support from districts as well as changes in the designs themselves as they seek to adapt to local contexts. In this report, we focus largely on between-school variance.1

**ORGANIZATION OF THE CHAPTER**

The remainder of this chapter is divided into several sections.

- The next section provides details about the core implementation index we use to measure implementation;
- We then discuss the challenges inherent in measuring implementation within and across designs, given that each design is unique and that designs themselves may be evolving over time and adapting to local environments;
- The following section presents findings from our analyses of the index, using the longitudinal sample; and
- A final section provides a brief summary of the chapter.

---

1In analyses not reported here, we found that declines in the variability of reported implementation both within and between these schools were small between 1997 and 1998; however, in almost every case, the variance within school increased between 1998 and 1999. It was also true that the between-schools variance in implementation was much smaller than the within-school variance, and was generally stable over time, with a couple of exceptions. These analyses were conducted using a different index than the one reported here. Our reviewers suggested omitting these analyses from this report for reasons of clarity. For further details, see Berends, Kirby, et al., 2001.
CONSTRUCTING A CORE IMPLEMENTATION INDEX²

The core implementation index is a summative scale of teacher responses as to the degree to which the following described their school (on a scale of 1–6, with 1 = does not describe my school, and 6 = clearly describes my school):³

- 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.

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

---
²In our earlier report, Berends, 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. We also measured implementation in the current report using this index, but because the analyses did not add substantially to the conclusions, we do not show them here.
³The alpha reliability of this index was 0.81. The range of correlations for the individual items was from 0.21 to 0.57.
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.

MEASURING IMPLEMENTATION WITHIN AND ACROSS DESIGNS: SOME CAVEATS

As we discussed in our earlier work, measuring progress in implementation broadly across a wide set of schools in several partnering jurisdictions involved a number of challenges (Berends, Kirby, et al., 2001).

First, each design is unique. Attempting to develop a common set of indicators that measures implementation across designs is difficult, particularly when design teams adapt their programs to the local needs of the schools (Bodilly, 2001). However, despite their differences, design teams do aim to change some key conditions of schools in common ways, such as school organization, expectations for student performance, professional development, instructional strategies, and parent involvement.4 We attempted to draw on these commonalities to guide the construction of an index that could be used to broadly measure “core” implementation across designs.

Second, the difficulties of constructing indices that capture the key components of a design are compounded by the fact that these design components may themselves be evolving (see Bodilly, 2001). For example, design teams may change their implementation strategies because of lessons learned during development and implementation experiences in various sites.

4With the recent support of the federal CSRD program, schools need to make sure that their plan covers these areas. If one particular design team or CSRD model does not cover these and several other areas of school improvement, then schools need to adopt more than one design or model (see Kirby, Berends, et al., in review).
Third, even if one developed measures on which there was general agreement that they fully captured the key facets of designs, the local context introduces a great deal of variability that must be taken into account (Bodilly, 1998; Bodilly and Berends, 1999). For example, while a design may focus on project-based learning over several weeks of the semester, this may be superseded by district-mandated curricula that take priority over significant portions of each school day.

Fourth, because the index is so general, it may be measuring more than just reform implementation. Each of the components is a characteristic of effective schools, so schools may be pursuing these separately as school goals or as part of a district initiative. An increase in any one of these measures may not necessarily mean higher implementation of the model. For example, it may be that the design is helping the school to better attain these goals, or even that the school has been more successful in meeting these goals over time, independent of the model.

Fifth, it is important to note that all the results reported here on implementation are based on teachers’ responses to surveys. The usefulness of what we can learn and infer from the analyses is heavily dependent on the quality of the data that are obtained from these surveys. In some instances, what we find has been validated by RAND’s early case studies and other research (Bodilly, 1998; Ross et al., 1997; Datnow and Stringfield, 1997; Stringfield and Datnow, 1998), but for some indicators, all we have are teacher-reported survey measures.

Sixth, 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, in the school-level descriptive analyses in this chapter, we focus on what appear to be educationally substantive differences where appropriate.

---

5We thank one of our reviewers, Amanda Datnow, for making this point.
Despite these challenges, as said earlier in Chapter 1, 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 the NAS schools during the scale-up phase.

Our focus in this chapter is on variation in implementation both among jurisdictions as well as design teams, where appropriate. The small sample size for some designs, the widely differing environments facing designs implemented in different jurisdictions, and the varying degree of support from districts make comparisons of implementation across design teams somewhat problematic.

In what follows, the school is the unit of analysis. In some instances, we conduct multiple comparison tests of differences among means, but, as we mentioned earlier, small sample sizes mean that finding statistically significant differences is somewhat more difficult in these data. In addition, outliers can make the means less representative than in more well-behaved distributions. These tests remain useful, however, in highlighting differences across jurisdictions and design teams that are both substantively and statistically meaningful. Nevertheless, our primary emphasis in this chapter is on the patterns that emerge from the analyses taken together rather than on particular differences for any given indicator.

---

6Statistically 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. When conducting multiple tests, say n, and setting the critical level to $\alpha$ for each test, the chances of falsely rejecting at least one of the hypotheses is $1 - (1 - 0.05)^n$ if the tests are independent. The Bonferroni correction allows us to control for the fact that we are conducting multiple tests and to ensure that the overall chance of falsely rejecting each hypothesis remains $\alpha$. 
RESULTS: DIFFERENCES IN IMPLEMENTATION BY JURISDICTION AND DESIGN TEAM

We began by analyzing differences in the mean implementation level in 1999 by jurisdiction and design team. We then focused on changes over time. 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.

Differences in Implementation by Jurisdiction, 1999

Our earlier work had found large differences in the distribution of the core implementation index across the jurisdictions as well as design teams. Figures 4.1 and 4.2 show the distribution of this normalized index by jurisdiction and design team for 1999, when all schools had been implementing for three or more years. Kentucky and Memphs ranked relatively high on this index with means that were 0.60 and 0.33 of a standard deviation higher than the overall mean while Washington and San Antonio ranked the lowest, with means that were 0.77 and 0.87 of a standard deviation lower than the overall mean. This ranking mirrors what we had found in 1998, with the exception that Philadelphia, which had ranked second lowest in terms of implementation, now ranks higher. Memphs schools also displayed the greatest spread in the data, as is evident from the long

---

7The first three graphs of this chapter 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.

8We thank our reviewer, Robert Croninger, for his suggestion that we use standardized metrics to make calculation of effect sizes easier for the reader.

9The 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 = 0.60.
# Implementation in a Longitudinal Sample of New American Schools

## Standardized z-Score

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cincinnati</td>
<td>n = 10</td>
</tr>
<tr>
<td>Dade</td>
<td>n = 1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>n = 13</td>
</tr>
<tr>
<td>Memphis</td>
<td>n = 24</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>n = 6</td>
</tr>
<tr>
<td>San Antonio</td>
<td>n = 7</td>
</tr>
<tr>
<td>Washington</td>
<td>n = 10</td>
</tr>
</tbody>
</table>

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

### Figure 4.1—Core Implementation Index (standardized z-scores) by Jurisdiction, Spring 1999

- 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.

### Differences in Implementation by Design Team, 1999

Comparisons among design teams reveal that CON, RW, and NARE ranked comparatively high on the core implementation index while MRSF generally ranked the lowest, reflecting the ranking we found in 1998 (Figure 4.2). CON schools had a mean that was almost one
standard deviation higher than the overall mean while RW and NARE schools had means that were 2/10ths and 1/10th 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, none of these differences in means was statistically significant.

The results for the core implementation index by jurisdictions and design teams are consistent with the results that we found when examining a wider set of indicators separately (see Berends, Kirby, et al. 2001, Appendix B). In addition, we also performed sensitivity analyses of our results by constructing more diverse and more inclusive indices of implementation (see Berends, 2000, for an example of a larger index). The results were consistent across these different indices.
RESULTS: CHANGES OVER TIME IN LEVELS OF IMPLEMENTATION

We measured change over time in several different ways:

- Changes in the overall mean implementation index across all schools;
- Changes in the mean implementation index by jurisdiction and design teams; and
- Changes in the overall mean implementation levels by number of years schools had been implementing in 1997 (one year, two years, three or more years) and changes in the components of the implementation index.

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 one.

Changes in Mean Level of Implementation, 1997–1999

Figure 4.3 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. 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.\(^{10}\)

The spread declined over time as well, as can be seen from the figure. Although not shown here, the variance in mean implementation

---

\(^{10}\)These are calculated as follows: The mean implementation index was 4.29 in 1998; thus the z-score for the 1998 mean is \((4.29–4.14)/0.61=0.25\). Similarly, the z-score for the 1999 mean is \((4.32–4.14)/0.61=0.29\).
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 not statistically significant. However, the within-school variance increased from 0.86 to 0.97 over the same time period, suggesting that implementation did not become more “schoolwide” within a school.

**Changes in Implementation Across Jurisdictions and Design**

Table 4.1 shows the mean implementation level and change in the mean over time for the jurisdictions and design teams in our sample. Among jurisdictions, the largest gains were posted by Philadelphia and Washington schools; these schools had among the lowest levels of implementation in 1997. Memphis schools gained about 3/10ths of a standard deviation and Kentucky about 2/10ths of a standard deviation, relative to their 1997 means.
Table 4.1

Core Implementation Index: 1997 Mean and Changes over Time
(in standard deviation units), by Jurisdiction and Design Team, 1997–1999

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Number</th>
<th>1997 Mean</th>
<th>1997 Standard Deviation</th>
<th>Change Between 1997 and 1999 (standard deviation units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cincinnati</td>
<td>10</td>
<td>4.31</td>
<td>0.59</td>
<td>0.15</td>
</tr>
<tr>
<td>Dade</td>
<td>1</td>
<td>3.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Kentucky</td>
<td>13</td>
<td>4.56</td>
<td>0.34</td>
<td>0.19</td>
</tr>
<tr>
<td>Memphis</td>
<td>24</td>
<td>4.56</td>
<td>0.52</td>
<td>0.31</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>6</td>
<td>3.47</td>
<td>0.33</td>
<td>1.80</td>
</tr>
<tr>
<td>San Antonio</td>
<td>7</td>
<td>3.79</td>
<td>0.79</td>
<td>0.10</td>
</tr>
<tr>
<td>Washington</td>
<td>10</td>
<td>3.69</td>
<td>0.40</td>
<td>0.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design team</th>
<th>Number</th>
<th>1997 Mean</th>
<th>1997 Standard Deviation</th>
<th>Change Between 1997 and 1999 (standard deviation units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>4</td>
<td>3.62</td>
<td>0.39</td>
<td>1.87</td>
</tr>
<tr>
<td>AT</td>
<td>13</td>
<td>3.71</td>
<td>0.36</td>
<td>1.13</td>
</tr>
<tr>
<td>CON</td>
<td>6</td>
<td>4.62</td>
<td>0.47</td>
<td>0.44</td>
</tr>
<tr>
<td>EL</td>
<td>10</td>
<td>4.27</td>
<td>0.61</td>
<td>–0.09</td>
</tr>
<tr>
<td>MRSIH</td>
<td>7</td>
<td>3.59</td>
<td>0.52</td>
<td>0.86</td>
</tr>
<tr>
<td>NARE</td>
<td>19</td>
<td>4.28</td>
<td>0.57</td>
<td>0.16</td>
</tr>
<tr>
<td>RW</td>
<td>12</td>
<td>4.52</td>
<td>0.54</td>
<td>–0.17</td>
</tr>
</tbody>
</table>

Overall mean 71 4.14 0.61 0.29

*aThe standardized change shown for each group was calculated using its own mean and standard deviation.

Among design teams, AC, AT, and MRSIH showed the largest gains; schools implementing these designs had much lower than average levels of implementation in 1997. CON schools that ranked the highest in implementation in 1997 gained close to half a standard deviation, while RW and EL both posted declines of between 1/10th and 2/10ths of a standard deviation relative to their 1997 means. While it is interesting to examine changes over time by design team, one needs to be cautious in drawing inferences from these data. The small sample sizes and differences in the composition of the groups with respect to the number of years schools have been implementing designs make these comparisons less precise than one would wish.
Changes in Implementation by Years of Implementation

In order to see whether implementation has deepened over time, it is important to disaggregate the change in the mean level of implementation by number of years that schools have been implementing.

Table 4.2 shows the means for the core implementation index and its components, for schools grouped by years of implementation in 1997—one year, two years, and three or more years. Note that we do not use a standardized metric here because each group of schools has a different mean and a different standard deviation for the base year (1997). As a result, we focus on actual changes in this table. Table C.1 (Appendix C) provides the standard deviations of the core implementation index and its components. In the discussion, we provide an indication of effect sizes, where appropriate.

Note, for the first group of schools, the change between 1997 and 1998 measures the change in reported implementation levels between year 1 and year 2 of implementation, and the change between 1998 and 1999 measures the change in reported implementation levels between year 2 and year 3. For the second group of schools, the change between 1997 and 1998 measures the change in reported implementation between year 2 and year 3 of implementation and the change between 1998 and 1999, the change in reported implementation between year 3 and year 4. Similarly, for the third set of schools, the change between 1997 and 1998 measures the change in reported implementation between year 3 and year 4, while the change between 1998 and 1999 measures the change in reported implementation between year 4 and year 5 of implementation.

Changes in the Mean Implementation Level. Implementation appears to increase and deepen over the first four years after schools adopt designs, although at a decreasing rate, lending support to our hypothesis that implementation is likely to be a polynomial function of time. For example, the core implementation index:
Table 4.2
Means of the Core Implementation Index and Its Components, by Number of Years of Implementation, 1997–1999

<table>
<thead>
<tr>
<th>Components of the Core Implementation Index</th>
<th>Schools that had been implementing for one year in 1997 (n = 31)</th>
<th>Schools that had been implementing for two years in 1997 (n = 22)</th>
<th>Schools that had been implementing for three or more years in 1997 (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents and community members are involved in the educational program</td>
<td>3.50</td>
<td>3.60</td>
<td>3.75</td>
</tr>
<tr>
<td>Student assessments are explicitly linked to academic standards</td>
<td>4.07</td>
<td>4.47</td>
<td>4.71</td>
</tr>
<tr>
<td>Teachers develop and monitor student progress with personalized, individualized learning programs</td>
<td>3.76</td>
<td>3.97</td>
<td>4.11</td>
</tr>
<tr>
<td>Student grouping is fluid, multiage, or multiyear</td>
<td>3.03</td>
<td>3.34</td>
<td>3.43</td>
</tr>
<tr>
<td>Teachers are continual learners and team members through professional development, common planning, and collaboration</td>
<td>4.01</td>
<td>4.76</td>
<td>4.83</td>
</tr>
<tr>
<td>Performance expectations are made explicit to students so that they can track their progress over time</td>
<td>3.91</td>
<td>4.27</td>
<td>4.29</td>
</tr>
<tr>
<td>Core implementation index</td>
<td>3.80</td>
<td>4.12</td>
<td>4.17</td>
</tr>
</tbody>
</table>
• Increased by 0.32 between year 1 and year 2 of implementation for those schools that reported only one year of implementation in 1997, a large change amounting to a little over half of a standard deviation;

• Increased by 0.05–0.09 between year 2 and year 3 of implementation (two groups of schools). This amounts to between $1/10^{th}$ and $2/10^{ths}$ of a standard deviation;

• Increased by 0.01–0.09 between year 3 and year 4 of implementation (two groups of schools) (between 0.02 and 0.18 of a standard deviation);

• Declined by –0.11 between year 4 and year 5 of implementation (almost $2/10^{ths}$ of a standard deviation). Almost all these schools were NARE and RW schools.

If we further disaggregate schools in the fifth year of implementation, we find that there were five schools that had more than five years of implementation (two schools reported six years, two reported seven, and one reported ten years of implementation). While these sample sizes are small, we should note that these schools did display higher levels of implementation than those in the fifth year. For example, mean implementation levels were 4.19 for the 13 schools in the fifth year, 4.86 for the two schools in the sixth year, 4.60 for the two schools in the seventh year, and 4.68 for the one school in the tenth year of implementation. Note that the decline in implementation between the fourth and fifth years is very large: almost half a standard deviation.

Figure 4.4 summarizes the actual relationship between years of implementation and the level of implementation for schools in our sample.11 We see a sharp increase between the first and second

---

11In 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.
year, 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.

Changes in the Components of the Core Implementation Index

Looking at the other indicators, we find:

- For schools in the first four years of implementation, every indicator that comprised the core implementation index increased between 1997 and 1999 (Table 4.2). Some showed particularly large changes. For example, the indicator “student assessments are explicitly linked to academic standards” showed an increase of 0.64 for schools in the first group and 0.25 for schools in the second group. This was the only indicator that did not show a
decline for schools beyond the fourth year of implementation. This is likely linked to standards-based reform being adopted in most states (see Berends, Kirby, et al., in review) and the new Title I requirements under which student assessments need to be explicitly linked to content and performance standards.

- As expected, the mean for every implementation indicator is higher in every case for schools that had been implementing for two or more years compared with schools implementing for one year. This suggests that implementation does deepen with maturity. However, for schools implementing for three or more years in 1997, the changes in indicators between 1997 and 1998 are quite small and negative in some instances, and negative in almost all instances between 1998 and 1999, suggesting that implementation of these components appears to have reached a plateau in these schools, and has started to decline. This deserves further attention to understand the reasons for the decline.

Changes in Between-School Variance of Implementation

The table in Appendix C (Table C.1) shows the standard deviations for the core implementation index and its components, for schools grouped by years of implementation in 1997—one year, two years, three or more years. We summarize some findings here. The overall pattern is not very clear:

- Overall, as we had seen earlier in Figure 4.3, the variance in reported levels of mean implementation across all schools declined over time by about 15 percent.

- The variance between schools in the group that had been implementing for one year in 1997 (i.e., those relatively early in the implementation process) declined markedly between 1997 and 1999 (by about 15 percent).

- However, variance between schools remained fairly stable for the other groups.

- The indicators with the greatest between-school variability are the ones related to student grouping and to parent/community involvement. Variance with respect to the latter increased
markedly over time, especially among schools that had been implementing for two or more years in 1997.

SUMMARY

The following are the main results from the updated analyses of implementation reported in this chapter.

- **Overall mean implementation across all schools increased modestly by about 3/10ths of a standard deviation.** Between-school variance declined somewhat between 1997 and 1999, but within-school variance increased.

- **There were large differences in implementation by jurisdiction.** Similar to what we found in our earlier study (Berends, Kirby, et al., 2001), Kentucky and Memphis ranked relatively high on this index, while Washington and San Antonio ranked the lowest.

- **There were differences in implementation by design 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 what we found in 1998.

- **Implementation appears to increase and deepen over the first four years after schools adopt designs, although at a decreasing rate.** Between the fourth and fifth year, however, we see a significant downturn in implementation, although a few schools showed increases beyond that time period.