Chapter Five

FACTORS AFFECTING IMPLEMENTATION

The conceptual framework outlined in Chapter Two pointed to the possible interrelationships between teacher background characteristics, school characteristics, district support, and assistance provided by design teams and implementation of the key components of designs. The previous chapter described a summary measure of implementation; however, as we noted, there is a great deal of variability in the overall indices across jurisdictions and design teams, as well as within schools. The multivariate analyses presented in this chapter allow us to analyze this variation further by partitioning it into within- and between-school components and relating this variation to a variety of teacher, school, and design team factors.[1]

VARIATION IN IMPLEMENTATION: MULTILEVEL ANALYSIS

Because the data are nested—that is, teachers are nested within schools—we use multilevel modeling techniques to provide more accurate estimates of the relationships between the dependent and independent measures (see Bryk and Raudenbush, 1992; Bryk, Raudenbush, and Congdon, 1996; Singer, 1998).

---

[1]Recall that our overall core index is comprised of the following: 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.

First, we partition the variance in the dependent variables into their within-school and between-school components. This provides information about whether most of the variance in the dependent variable lies between schools or within schools. Next, we estimate a set of regression coefficients in each school (Level 1), and then the constant term in the first equation (Level 1) becomes an outcome to be explained by school demographic and implementation factors (Level 2).

Specifically, for the implementation index, we estimate the following models:

Individual Teacher Model (Level 1)

$$Y_{ij} = \beta_{0j} + \beta_{\iota} X_{ij} + r_{ij}$$

where

- Y_{ij} is the dependent variable, the teacher-reported implementation index;
- β_{0j} is the constant term, and it is the average value of the dependent variable in the school j;
- β_i is a vector of the Level 1 coefficients to be estimated with school j for the listed independent variables;
- X_{ij} is a vector of independent variables measured at the teacher level; and
- r_{ij} is the Level 1 random effect, assumed to be normally distributed with a mean of zero and constant variance.

School Context Model (Level 2)

$$\beta_{0j} = \gamma_0 + \gamma_j Z_j + u$$

where

- β_{0j} in this model is from the teacher level equation above;

- γ_0 is the constant term;
- γ_j is a vector of the Level 2 coefficients for the independent school-level variables;
- Z_j is a vector of the independent variables measured at the school level; and
- u is a Level 2 random effect.

The multilevel models described above are simple "fixed coefficient" models (Kreft and DeLeeuw, 1998). In other words, the coefficients estimating the Level 1 relationships between teacher characteristics and implementation are held constant across schools. There are no cross-level interactions between teacher and school characteristics. Thus, between-school differences are limited to differences in intercepts. In other words, the intercept for each school is the sum of the overall intercept and the sums of the school aggregate variables weighted by the school-level regression coefficients, plus error. The implementation index of each teacher is then the sum of that teacher's school intercept and the sum of the teacher-level background variables weighted by the teacher-level regression coefficients, plus error (Koretz, McCaffrey, and Sullivan, 2001).

Caveat

It is important to remember that our sample of NAS schools is not a random sample, but a sample of all those schools in the seven partnering jurisdictions that reported implementation during 1997–1999 and had survey information from teachers and principals. Despite this not being a random sample, we use multilevel modeling techniques to explore the relationships among variables at one point in time—spring 1999. These provide a more accurate description of the relationships in these NAS schools for this particular point in time. We estimate the model using the longitudinal sample and examine whether factors affecting implementation changed over time, from 1998 to 1999. We then estimate the model on the combined sample and formally test for differences between the longitudinal and the combined samples.

OPERATIONALIZING THE VARIABLES AFFECTING IMPLEMENTATION

The factors listed below are drawn from our conceptual framework. We describe the operationalization of each variable before showing its relationship with implementation in a multivariate framework.

Teacher Characteristics

Dummy variables for race-ethnicity (African American, Latino, other, vs. non-Hispanic white), educational degree (master's vs. bachelor's), total teaching experience (measured in years), and age (30 years or older vs. less than 30).

In addition, we also include three measures of teacher perceptions of their students' ability and readiness to learn (that to some degree measure teacher efficacy) and a variable that measures teacher perceptions of principal leadership:

- *Lack of basic skills is a hindrance to your students' academic success:* Teachers ranked this on a scale of 1–4 with 1 = great hindrance and 4 = not at all.
- *Lack of student discipline and inadequate parent support for students is a hindrance to your students' academic success:* These two separate survey items were summed to create one index. Each was ranked on a scale of 1–4 with 1 = great hindrance and 4 = not at all. We combined the two and obtained an average value.
- *Most of my students can learn with the school resources available to them:* This was measured on a four-point scale with 1 = strongly disagree and 4 = strongly agree.
- *Principal leadership:* Teachers were asked several questions regarding the degree of support and leadership provided by the principal. These included communicating clearly what is expected of teachers, supportive and encouraging behavior, getting resources for the school, enforcing rules for student conduct, talking with teachers regarding instructional practices, having

confidence in the expertise of the teachers, and taking a personal interest in the professional development of teachers. We combined these into a summative index of principal leadership; the alpha reliability for this index was 0.87, and the correlations ranged from 0.28 to 0.65. Unfortunately, these data were not available for Dade schools because we had to administer an attenuated survey in these schools.

School Characteristics

- *Large school:* This is a dummy variable equal to one if the school is large (i.e., 400 students or more), and equal to zero otherwise.
- *Secondary school:* This is a dummy variable equal to one if the school is secondary, and equal to zero otherwise.
- *Percent student mobility:* This is a continuous variable measuring the overall student mobility reported by the principal.
- *School poverty:* This is a continuous variable measuring the percentage of students receiving free/reduced price lunch.
- *School minority composition:* This is a continuous variable measuring the percentage of students who are non-white.
- *Interaction between poverty and minority composition:* This is a continuous variable measuring the interaction between the two variables.
- *Principal leadership:* We created a school-level mean based on the sample of teachers in each school that measured principal leadership at the school level.
- *Implementation index 1997:* Because we are focusing on change and have repeated measures of implementation, we introduce the school implementation level in 1997 as a control variable. While this is not an "explanatory" variable, it makes use of the longitudinal nature of the data and allows us to see how progress in implementation is related to prior levels of implementation.

Designs and Design Team Assistance

Teacher Level

- *Communication by designs to schools:* measures the degree to which individual teachers report that the design team clearly communicated the design so that it could be well implemented. Scores for this variable range from 1 = did not communicate clearly at all to 6 = definitely communicated clearly.
- *Teacher support:* measures the extent to which teachers reported supporting or opposing the design in their school. This variable was measured on a 5-point scale, where 1 = strongly oppose and 5 = strongly support.

School Level[2]

- *Years implementing:* is number of years the school has been implementing the design with a range from two to four or more (from principal interview).
- *Communication by designs to schools:* is a school-level mean created from individual teacher reports regarding level of communication by design teams.
- *Teacher support:* is a school-level mean created from individual teacher reports; this captures the aggregated level of support in a school.
- *Set of dummy variables for designs:* Because there may be unmeasured characteristics of designs that may influence implementation in addition to those mentioned above, we also included in some models a set of dummy variables that were equal to one if the school had adopted that particular design, zero oth-

---

[2]In analyses of earlier data, we had included a dummy variable equal to one if 60 percent or more of teachers actually voted to implement the design, zero otherwise. The models reported here did not include this variable for two reasons. First, it measured support three years ago when the schools were considering the design and we had more direct measures of support. Second, given turnover, this variable may not correctly reflect level of support in the spring of 1999.

erwise (RW is the reference category because of its relatively high levels of implementation discussed earlier).

In addition to these variables, we also created an index that measured the availability of resources at the school level to implement the designs.[3] This variable was correlated with principal leadership and was statistically insignificant when both variables were included in the model. In the interest of parsimony, it was omitted from the final model.

District Support

The models include one of the two measures of district support:

- *Index of district support:* We ranked jurisdictions along a variety of key dimensions based on Bodilly's earlier work (1998) as well as additional interviews with districts in 1998. As we mentioned earlier, Bodilly identified several district and institutional factors that contributed to implementation: leadership backing and stability; centrality of effort; lack of crises; history of trust and cooperation; resource support; school autonomy; union support; and aligned assessment. Jurisdictions were ranked on these dimensions; the various rankings were then summed into an overall summative index of district support. The alpha reliability

[3]*Resources Index* was a school measure based on aggregated teacher reports about whether the school had sufficient resources to implement the designs. Teachers could respond to several questions using a 5-point scale ranging from "no resources are available" to "all are available." The resources index is a combination of a number of questions that asked the extent to which the teacher's school had the resources needed to implement the major elements of the design. These questions included:

- Materials to describe the program;
- Materials to support instruction;
- Professional development for teachers;
- Time for planning, collaboration, and development;
- Staff or consultant to mentor, advise, and provide ongoing support;
- Technology; and
- Funds and funding flexibility.

The alpha reliability of this scale was 0.92 (correlations among these items ranged from 0.50 to .090).

of this index was 0.85, and the correlations ranged from 0.16 to 0.77.

- *Set of dummy variables for each jurisdiction:* Because the above index is open to criticism both on the grounds of being largely judgmental as well as being dated (based on data collected during the 1996–1997 school year), we also used the set of separate dummy variables for each jurisdiction, which was used in the earlier chapter, to test in the multilevel analyses (Memphis is the reference category). Another reason for using this set of dummy indicators rather than the index is to capture dimensions of district support that have perhaps not been captured or only partially captured by the district index.

THE ANALYSIS SAMPLE

Table 5.1 provides the means and standard deviations (SDs) for the variables in the models. The implementation index which had a mean of around 4.3 on a 1–6 scale was standardized to a mean of zero and SD = 1 in the estimated models. The analysis sample consists of about 1,200 teachers and 70 schools.[4]

Teacher Characteristics

About 63 percent of the teachers were non-Hispanic white, a little over one-fifth were African American, and about 3 percent were Hispanic. Less than 2 percent of the sample reported their race/ethnicity as "other," and about 12 percent were missing data on race/ethnicity. These latter two categories were combined into an "other/missing" category. The remaining 60–62 percent were non-Hispanic white. About half the sample had a master's degree or higher and over two-thirds were 30 years or older (a good proportion were 40 or older). They had been teaching about 16 years on average.

A large proportion of the sample felt that lack of basic skills was a hindrance to their students' academic success. This was reflected in

---

[4]Note that the sample of schools is 70, not 71. We dropped the one school in Dade for reasons explained later in this section.

Table 5.1

Descriptive Statistics of Variables in Multilevel Analysis

Variables	Mean	SD
Dependent variable		
Implementation index (range 1–6)	4.30	1.08
Implementation index (standardized)	0.00	1.00
Independent variables for teachers (n = 1092)[a]		
Race-ethnicity		
African American	0.20	
Hispanic/Latino	0.03	
Other/Missing	0.14	
Educational degree: master's or above	0.54	
Age: ≥ 30 years	0.69	
Years of teaching experience	16.57	11.45
Lack of basic skills a hindrance (range 1–4, 1 = great hindrance, 4 = not at all)	1.85	0.88
Student discipline and inadequate parent support a hindrance (range 2–8, 2 = great hindrance, 8 = not at all)	4.70	1.80
Students can learn with available resources (range 1–4, 1 = strongly disagree, 4 = strongly agree)	3.27	0.80
Principal leadership index (range 1–4, 1 = low, 4 = high)	3.23	0.65
Design-related variables		
Communication by designs to schools (range 1–6, 1 = not at all, 6 = yes, definitely)	3.88	1.67
Support for design (range 1–5, 1 = strongly oppose, 5 = strongly support)	3.59	1.23
Independent variables for schools (n = 70)[b]		
Percent poverty in 1998	66.29	29.39
Percent minority in 1997	61.52	37.42
Poverty x minority	48.99	38.57
Implementation index 1997	4.15	0.61
Years implementing	3.97	1.24
Principal leadership index (range 1–4, 1 = low, 4 = high)	3.21	0.39
Design-related variables		
Communication by designs to schools (range 1–6, 1 = not at all, 6 = yes, definitely)	3.90	1.08
Support for design (range 1–5, 1 = strongly opposed, 5 = strongly support)	3.65	0.59
Jurisdictions:		
Cincinnati	0.14	
Kentucky	0.19	
Philadelphia	0.09	
San Antonio	0.10	
Washington	0.14	
Index of district support in 1997 (range 9–24, 9 = not supportive, 24 = very supportive)	17.63	2.53
Design teams:		
AC	0.043	
AT	0.186	
CON	0.086	
EL	0.143	
MRSH	0.100	
NARE	0.271	
RW	0.171	

[a]Standard deviations are calculated from the teacher sample.

[b]Standard deviations are calculated from the school sample.

the very low mean of around 1.8 for the sample. In terms of lack of student discipline and inadequate parent support, the teachers appeared to be at the midpoint on this measure with a mean of 4.6–4.7 on a scale that ranged from 2 to 8. Teachers were surprisingly positive about their students' ability to learn, given the available resources. The mean was quite high (approximately 3.3) on a scale of 1–4.

When considering whether design teams clearly communicated the design to school staff so that it could be well implemented, we found that the average score for this teacher-level measure was about 3.9–4.0, which was somewhat higher than the mid-range on a 6-point scale ranging from not at all clear (1) to definitely clear (6). There was a great deal of variation around this mean as indicated by the standard deviation of over 1.6. In terms of support for the design, teachers again were in the middle with a mean of 3.6–3.7 (SD = 1.2).

In 1998, about 70 percent of teachers had attended a design team workshop in the past twelve months; not surprisingly, this was much lower in 1999.

School Characteristics

Poverty rate was measured as of 1998 and minority composition as of 1997. The mean poverty rate for the sample was 66 percent and the standard deviation was quite large, 29 percent. The schools also had high proportions of minority students: the mean for the sample was a little over 60 percent, with a standard deviation of 37 percent.

Given the constraints of our sample size (70 schools), we estimated several versions of the model with combinations of school characteristics. School size, school level, and student mobility were consistently insignificant in these models with small estimated effects. As a result, in the interests of parsimony and statistical power, we excluded these from the models shown here.

On average, schools were in their fourth year of implementation by 1999.

The mean school-level implementation index for 1997 was 4.15 with a standard deviation of 0.61. Note that the variability of this school-level index is much smaller than the teacher-level implementation index, the dependent variable in the model, because of aggregation at the school level.

The mean for the principal leadership index was relatively high: 3.2–3.3 on a scale of 1–4 and the standard deviation was 0.4. The schools ranged from 1.9 to 3.8 on this index. Recall that we do not have data on this variable for Dade schools. Given the importance of this variable (as shown in our earlier report, Berends, Kirby, et al., 2001), and given that we only had one school in Dade, we decided to drop this school from our multivariate analyses.

The school-level mean for communication by the design team was 3.90 on a scale of 1–6, with a large standard deviation of 1.08. The school-level mean for teacher support was 3.65 on a scale of 1–5.

The set of dummy indicators for jurisdictions shows the proportion of schools in our sample that were in particular jurisdictions. For example, 14 percent of our schools were in Cincinnati while 19 percent were in Kentucky. The largest numbers of schools in our sample were in Cincinnati, Kentucky, and Washington.

The index of district support was about 18 on a scale of 9–24.

The set of dummy variables for the design teams shows that AC schools were somewhat less well-represented in our sample, accounting for less than 4 percent of the sample. CON and MRSH schools comprised between 9 and 10 percent of the sample. The largest group was NARE schools, which constituted 27 percent of the sample, while other designs—AT, EL, RW—each accounted for between 14 and 19 percent of the sample.

MULTIVARIATE RESULTS[5]

In order to facilitate interpretation of the intercept term in the models, we centered all variables that were unique to Level 1 or Level 2 at their grand means. Centering is particularly important for the Level 1 variables both for interpretation of the intercept as well as to ensure

---

[5] We are grateful to our reviewer, Robert Croninger, for several helpful suggestions regarding the multilevel models presented here. This section benefited greatly from his constructive advice.

numerical stability (Bryk and Raudenbush, 1992: 25). Without centering, one would interpret the intercept as the expected outcome for a teacher in school j who has a value of zero on all the Level 1 predictor variables, which often does not have meaning. With centering, one can interpret the intercept as the expected outcome for a teacher in school j who is at the mean of all the predictor variables. Dummy variables were also grand-mean centered; thus the intercept term is the adjusted mean outcome in school j, adjusted for differences among units in the percentages of teachers with various characteristics. Centering the Level 2 variables simply adds to the convenience, so that the intercept term can be interpreted as the expected outcome for a school that is at the mean of the sample in terms of school characteristics.

The variables that were included at both levels (communication by the design team; teacher support; and principal's leadership) were group-mean centered at the teacher level (by subtracting the school mean score from each individual teacher's score) and the school means were then entered at Level 2 into the model. As Bryk and Raudenbush (1992) point out, this procedure avoids the assumption that the effects of variations in school means equals the effects of deviations within schools and makes the model's coefficient straightforward estimates of the within- and between-school effects.[6]

Variance Components of the Dependent Variable

Before examining the relationship between implementation and teacher and school characteristics, we begin by partitioning the variance in the dependent measure into its between- and within-school components, and examining the reliability of each school's sample mean as an estimate of its true population mean.

---

[6]Cohen, Baldi, and Rathbun (1998) point out the importance of reintroducing the school means when the Level 1 variables are group-mean centered: "A model in which the (Level 1) variables are centered around their school means sheds an important piece of information: the school mean of the variable. When the analyst fails to re-introduce this source of systematic variation appropriately elsewhere in the model, he or she posits that the actual value of the centered variable does not influence the outcome, only the relative value (that is, relative to the school mean)" (1998: 18–19).

In 1999, we find that most of the variation in the dependent variable is at the teacher level, although a substantial proportion is between schools. For example, the estimated intraclass correlation, $\hat{\rho}$, that measures the proportion of the variance in the dependent variable between schools, is 0.18. Thus between-school variance accounts for 18 percent of the total variance in the dependent variable; the within-school variance accounts for the remaining 82 percent. The between-school variance component has declined from 1998, when it accounted for 27 percent of total variance, with a corresponding increase in the within-school variance.

Such findings are not uncommon in analyses of school contextual effects on student and teacher outcomes (see Lee and Bryk, 1989; Gamoran, 1992; Berends, 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.

We can also derive an estimate of the reliability of the sample mean in any school j, $\hat{\lambda}_j$, by substituting the estimated variance components into the following equation (Bryk and Raudenbush, 1992: 63):

$$\hat{\lambda}_j = \text{Reliability (School Mean}_j) =$$

$$\{\text{Estimated Between-School Variance}/[\text{Estimated Between-School Variance} + (\text{Estimated Within-School Variance}/\text{Sample Size}_j)]\}$$

The reliability of the sample mean as an estimate of the true school mean varies from school to school, depending on sample size. However, an overall measure of reliability can be obtained by averaging across the school reliabilities. For our data, $\hat{\lambda}_j = 0.74$, which, while not as high as one would prefer, indicates that the sample means tend to be fairly reliable as indicators of the true school means.

Multivariate Results

We estimated three different models to examine the relationships between the core implementation index and teacher, design team, school, and district factors:

- Model 1: includes teacher, school, and design team characteristics (excluding the design team dummy variables);
- Model 2: includes the full set of independent variables included in Model 1, plus jurisdiction dummy variables;
- Model 3: includes the full set of independent variables included in Model 1, plus design team dummy variables and an index of district support.

As mentioned earlier, in the multilevel regression models, the intercept is modeled as a random parameter, allowed to vary between schools. The teacher variables are included as fixed effects.

The results from the models are provided in Table 5.2.

Teacher-Level Effects

The strongest effect at the teacher level was teacher perceptions of principal leadership. Teachers who ranked one point higher on the principal leadership scale relative to the school mean reported implementation levels that were over half a standard deviation higher than teachers at the mean. This effect is statistically significant and consistent across all three models.

We found some interesting differences in the level of implementation reported by African American and Hispanic teachers vs. white, non-Hispanic teachers. African American teachers tended to report significantly higher levels of implementation than white teachers (about 0.20 and 0.24 of a standard deviation higher). The coefficient for Hispanic teachers is statistically significant only in Model 2, which controls for the jurisdiction dummy variables.

Among design-related variables, both clear communication and teacher support for the design were statistically significant. Teachers who were one point higher than the school mean on both these indicators reported implementation levels that were between 0.08 and 0.12 of a standard deviation higher. These effects were consistent across the models.

The variables measuring teacher perceptions about students' readiness and ability to learn were correlated with implementation. For

Table 5.2

Multilevel Results for the Relationships of Implementation to Teacher-, School-, and Design-Related Factors

	Model 1		Model 2		Model 3	
Variable	Coef-ficient	Standard Error	Coef-ficient	Standard Error	Coef-ficient	Standard Error
Intercept	–2.677**	0.390	–2.424**	0.340	–2.683**	0.349
Independent variables for teachers						
Race-ethnicity						
African American	0.245**	0.067	0.210**	0.067	0.241**	0.067
Hispanic/Latino	0.098	0.147	0.319*	0.160	0.142	0.146
Other/missing	–0.039	0.082	–0.062	0.081	–0.046	0.082
Educational degree: master's or above	–0.086	0.053	–0.102*	0.051	–0.091	0.051
Age: ≥ 30 years	0.085	0.081	0.105	0.081	0.081	0.081
Years of teaching experience	–0.002	0.002	–0.002	0.002	–0.002	0.002
Lack of basic skills a hindrance	0.075*	0.032	0.085**	0.031	0.083**	0.031
Student discipline and inadequate parent support a hindrance	0.023	0.016	0.023	0.016	0.022	0.016
Students can learn with available resources	0.069*	0.031	0.074*	0.031	0.073*	0.031
Principal leadership	0.576**	0.049	0.572**	0.049	0.578**	0.049
Design-related variables						
Communication by designs to schools	0.122**	0.019	0.121**	0.019	0.121**	0.019
Support for design	0.077**	0.024	0.075**	0.024	0.076**	0.024

Table 5.2 (continued)

Variable	Model 1 Coef-ficient	Model 1 Standard Error	Model 2 Coef-ficient	Model 2 Standard Error	Model 3 Coef-ficient	Model 3 Standard Error
Independent variables for schools						
Poverty x minority	–0.019**	0.005	–0.006	0.005	–0.017**	0.004
Poverty	0.012**	0.003	0.004	0.003	0.012**	0.003
Minority	0.011**	0.004	0.002	0.004	0.009**	0.003
Implementation index 1997	0.223**	0.082	0.159*	0.075	0.186*	0.081
Years implementing	0.026	0.037	–0.024	0.034	0.018	0.035
Principal leadership	0.586**	0.101	0.575**	0.090	0.555**	0.096
Design-related variables						
Communication by designs to schools	–0.005	0.046	–0.061	0.047	0.023	0.051
Support for design	0.223**	0.077	0.262**	0.068	0.215**	0.073
Jurisdiction						
Cincinnati	—a	—a	0.028	0.105	—a	—a
Kentucky	—a	—a	–0.077	0.135	—a	—a
Philadelphia	—a	—a	–0.159	0.121	—a	—a
San Antonio	—a	—a	–0.613**	0.134	—a	—a
Washington	—a	—a	–0.498**	0.178	—a	—a
Index of district support	—a	—a	—a	—a	0.001	0.015
Design teams:						
AC	—a	—a	—a	—a	0.312*	0.144
AT	—a	—a	—a	—a	0.001	0.114
CON	—a	—a	—a	—a	0.287*	0.130
EL	—a	—a	—a	—a	–0.129	0.117
MRSH	—a	—a	—a	—a	–0.223	0.127
NARE	—a	—a	—a	—a	0.089	0.155
Sample size						
Teachers	1,092		1,092		1,092	
Schools	70		70		70	

*Significant at .05 level.
**Significant at .01 level.
[a]Excluded from the model.

example, teachers who were one point higher than the mean with respect to their perceptions of student readiness and ability to learn were a little less than 1/10th of a standard deviation higher on the implementation index. This reinforces earlier literature on the importance of teachers' sense of efficacy in implementation (Fullan, 1991).

Mirroring what we had found earlier, teacher age, experience, and education did not appear to be important factors in implementation, controlling for everything else. This is similar to what Datnow and Castellano (2000) found in their study, although different from what Huberman (1989) reported. Teachers with a master's degree tended to report lower levels of implementation, although this is statistically significant only in Model 2.

School-Level Effects

We had chosen to enter three variables at both the teacher level and the school level, in order to be able to directly decompose the relationship between these variables and implementation into its within- and between-school components. These variables were principal leadership, communication by design team, and teacher support for the design. We discuss these first before commenting on the other school-level variables.

It is interesting that, at the school level, the effect of aggregated principal leadership is the largest predictor of between-school variance in implementation. A school that was one point higher on the school-level principal leadership scale reported implementation that was over half a standard deviation higher than a school at the mean of the index.

Teacher support was also important at the school level; schools that were one point higher than the sample mean in terms of teacher support were also likely to have implementation levels that were over 2/10ths of a standard deviation higher than schools at the mean. The effect was almost three times as large as at the teacher level.

Communication by the design team aggregated at the school level was insignificant in all three models.

An interesting question that arises in models like these is to what extent compositional or contextual effects are present. Such effects are said to occur when an aggregated school mean of a person-level characteristic has an effect on the outcome, even after controlling for the effect of the individual characteristic (Bryk and Raudenbush, 1992: 121–122). Note that in the case of principal leadership, there is no compositional effect present because the school-level effect is equal to the person-level effect. However, we do find a sizable compositional effect with respect to teacher support of approximately 0.14–0.18 of a standard deviation, even larger than the individual person effect.

As expected, the effect of prior implementation level was significant and tended to be dependent on whether jurisdiction or design team fixed effects were included in the model. Schools that ranked one point higher than average on the 1997 implementation index were likely to have implementation levels in 1999 that were about 0.15–0.22 of a standard deviation higher.

Some school demographics were significantly related to implementation, notably poverty and minority composition, both of which have positive effects on implementation in the two models not controlling for jurisdiction effects.[7] Schools that ranked 10 percentage points above the sample mean on either of these variables reported levels of implementation that were 1/10th of a standard deviation higher than schools at the sample mean. It is interesting and promising to find that schools serving largely poor or minority students report 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 indica-

---

[7]As mentioned earlier, we estimated several different versions of the models that included school size, school level, and student mobility. These variables were largely insignificant and were dominated by poverty and minority composition and were excluded from the final versions shown here. This is a change from our earlier results (Berends, Kirby, et al., 2001) that were based on a sample of 104 schools and 1998 data. In those models, school size, level, and student mobility did have significant effects on implementation. One plausible hypothesis for the change might be that as schools mature and gain experience with the designs, these variables become less important than variables describing the composition of the student body. Alternately, it could be that nonresponse in our panel data is leading to somewhat skewed results.

tion of the ability of some designs to help change these challenging schools.

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 is equal to the coefficients on poverty and minority. On net, taking the combined effect of poverty and minority composition, the effect of this variable largely washes out.

There was little difference in reported implementation by number of years schools had been implementing, net of the other variables.[8]

In Model 2, controlling for other variables, we find that among the jurisdiction dummy variables, the only two statistically significant variables are those for San Antonio and Washington, both of which rank significantly lower than Memphis, the reference category.

San Antonio schools are about 6/10^{ths} of a standard deviation lower on the implementation scale than schools in Memphis (the omitted category), after controlling for other factors. This is not surprising. Concerned about high-stakes accountability testing, in addition to the NAS designs, the district office introduced research-based instructional programs that targeted the two most basic subjects, reading and mathematics in all schools. Thus, by 1997–1998, schools were spending a substantial portion of the day on district-mandated curricular programs, rather than on design team activities. Within this context, it is not surprising that our surveys reveal very low levels of implementation of NAS designs in San Antonio.

Washington schools ranked about half a standard deviation lower than Memphis schools on the implementation index. Washington tended to rank low on indicators of district support.

These results are somewhat different from those reported in our earlier study (Berends, Kirby, et al., 2001), in which the model with jurisdiction dummy variables was dominated by the jurisdiction ef-

---

[8]A squared term to allow for nonlinearities in the relationship between years implementing and implementation was not significant in the model.

fects. These were all fairly sizable and mostly significant, even after controlling for school and teacher characteristics. The differences in results are due to the differences in the model specification and the sample. It is plausible to suggest that over time, in jurisdictions with higher implementation (Cincinnati, Memphis, Kentucky), maturity and experience with the designs ensures that the designs become institutionalized within schools, and schools are perhaps less vulnerable to, or in need of, direct district support and leadership. On the other hand, if schools that failed to receive district support also dropped out of the panel, then the district effects might well be underestimated. Based on the two studies and the RAND case studies, it is fair to infer that district and institutional factors are extremely important in ensuring the success of comprehensive school reform, particularly in the early stages of implementing the designs. This points to the need for a careful strategy to involve districts in any federal reform efforts such as CSRD.

We had earlier mentioned that we were interested in seeing whether characteristics of designs other than those measured in our models (clear communication, teacher support, attendance at design team workshop, adequacy of resources, the latter two of which are not included in models reported here) had an independent effect on implementation. We tested explicitly for these effects in Model 3. Controlling for prior implementation levels, and other factors, we find that compared with RW, the reference category, both AC and CON schools reported implementation levels that were about 3/10^{ths} of a standard deviation higher. AC schools have made considerable progress in implementation in the last year; they tended to be at the sample mean in earlier years, and lower than RW schools. The design team had been put on notice by NAS that they might be dropped from NAS's portfolio unless they improved. As a consequence, AC reduced the number of schools in which they were during scale-up and expended time and effort on the remaining schools. It clearly paid dividends, suggesting the importance of quality support, particularly if the design has limited capacity.

In our earlier study, design team differences were statistically insignificant, controlling for other factors. It may be that over time, unobserved differences in design teams become more important making some designs easier or harder to implement.

In Model 3, we also included a measure of district support based on a summary measure from RAND's case study analysis of these districts (Bodilly, 1998). The effect of this variable was small and statistically insignificant, although positive. One reason may be the variable is too dated—given that it was based on 1996–1997 data—to reflect current realities in terms of district environment and support.

Goodness of Fit

The models we estimated explained almost all of the between-school variance in implementation. As we had mentioned earlier, 18 percent of the variance in implementation in 1999 lies between schools; school-level measures considered here explain almost all of this between-school variance (Table 5.3). In fact, the test for the residual variance components for intercepts failed to reject the null hypothesis that the conditional component of the variance between schools was zero (Singer, 1998). Another way of examining this is to compute the residual intraclass correlation, the intraclass correlation among schools that are similar across the set of characteristics included in the models.[9] In the three models, this ranges from 2 to 5 percent.

Table 5.3

Variance in Implementation Explained by the Model

	Model 1	Model 2	Model 3
Sample			
Variance between schools (t)	0.183	0.183	0.183
Variance within schools (s^2)	0.818	0.818	0.818
Model			
Variance between schools (t)	0.030	0.009	0.011
Variance within schools (s^2)	0.563	0.564	0.565
Proportion of variance between schools explained by the model	84%	95%	94%
Proportion of variance within schools explained by the model	31%	31%	31%

[9]Recall that intraclass correlation is defined as that fraction of the sum of both variance components that occurs at the school level.

Explaining within-school variance has traditionally been much more difficult. The teacher-level factors included in the models explained about 31 percent of the within-school variance, suggesting that other measures are important for explaining the differences in implementation occurring within schools.

SUMMARY

The analyses shown here further our understanding about the progress NAS schools have made in the first several years of implementation. We find that teacher reports about implementation in their school differed much more within schools than between them.

Our models strongly underscore the importance of the following factors in implementation:

- Strong principal leadership;
- Teacher support for designs and clear communication;
- A sense of teacher efficacy in terms of positive teacher perceptions and expectations about students' ability and willingness to learn;
- Teacher and school demographic characteristics (e.g., African American teacher, schools serving large numbers of poor or minority students);
- Strong district support;
- Type of design; and
- Prior level of implementation.

Many of these factors have important policy implications for the success of the current CSRD reform effort.