CHAPTER 6:

THE PROMISE AND CHALLENGE OF EVALUATING SYSTEMIC REFORM IN AN URBAN DISTRICT

Thomas Corcoran
Consortium for Policy Research in Education

Ellen Foley
Annenberg Institute for School Reform

Philadelphia's Children Achieving Challenge

IN PHILADELPHIA, the Annenberg Challenge initiative, known as Children Achieving, set the ambitious goal of having every student achieve proficiency in three core subject areas – math, reading, and science – by 2008. With the support of $50 million from the Annenberg Foundation and a $100-million double match provided by Philadelphia businesses and local foundations, the School District of Philadelphia set out to achieve this goal by designing and implementing ambitious reforms in almost all aspects of its work, and, in the words of its fervent superintendent, David Hornbeck, to do it "all at once."

In this chapter, we describe what happened in Philadelphia in the late 1990s and what we learned about school reform during five years of studying the Children Achieving initiative. Drawing on both quantitative and qualitative data collected during the evaluation of Children Achieving, we examine its theory of action, its implementation, its successes and disappointments.[1] We also describe our experience

as evaluators, observing the initiative at close range and providing formative feedback. We argue that the initiative demonstrated some promising early gains in achievement. However, serious flaws in design and execution and inadequate attention to the Philadelphia context ultimately limited its impact and brought it to an end. In particular, we conclude that the policy dictum that everything had to be done at once, as well as poor sequencing of actions, failure to win teacher support for the reforms, and the emphasis on raising standardized-test scores led to uneven, often superficial, implementation. And we conclude that these flaws affected the evaluation by curtailing opportunities to provide candid, timely recommendations for midcourse corrections.

1. In 1996, the Children Achieving Challenge commissioned the Consortium for Policy Research in Education (CPRE) and Research for Action (RFA) to conduct a four-year evaluation of Philadelphia's Children Achieving initiative. CPRE conducted two systemwide surveys of teachers about the impact of the reforms on their daily work and about the character of their instruction. CPRE and RFA staff members also collected data from forty-eight Philadelphia schools by observing classrooms, meetings, and professional development sessions, and by interviewing teachers, principals, and other school officials. We interviewed district officials and civic leaders, and observed numerous meetings in which the reforms were debated, designed, and revised. We examined the SAT-9 test results and other indicators of system performance.

THE PLAN FOR TRANSFORMING PHILADELPHIA'S SCHOOLS

In 1995, with the support of the Annenberg Foundation, newly appointed Superintendent David Hornbeck launched Children Achieving, a ten-point reform agenda that promised what "no city with any significant number and diversity of students" had ever done before. Hornbeck boldly claimed that implementation of his plan would help "a large proportion of its young people achieve at high levels" (School District of Philadelphia 1995, p. i). The task was daunting. A special section of the *Philadelphia Inquirer* (1994) published just a few months earlier had painted a dismal portrait of the conditions in the school system. According to the *Inquirer*:

- Half the district's 220,000 students were from families on welfare.

- 136 of 238 schools were severely segregated.

- Over half of the city's public school students were failing to master basic skills. Fifty-one percent had failed the state reading test as compared to 13 percent statewide, and 50 percent failed the state math test as compared to 14 percent statewide. Seventy percent of African Americans and 75 percent of Latinos failed one or both parts of the state test.

- Forty-nine percent of ninth-graders failed to earn promotion to the tenth grade.

- On any given day one in four students was absent from class, and, in the average year, nearly one in four students was suspended from school.

To change these conditions and raise achievement, Hornbeck proposed an ambitious plan modeled after the reforms he had helped design for the state of Kentucky a few years earlier. The goals of Children Achieving were to transform the district into a school system characterized by

- high standards for all students

- accountability for results at all levels

- decentralization of authority from central office to schools and clusters of schools

- expanded professional development for teachers and administrators

- early childhood education for all children

- effective use of community services and supports

- adequate technology, instructional materials, and facilities

- strong public engagement

- adequate and effective use of resources

- comprehensiveness, or "Do all of the above at once"

The theory of action – the chain of logic about how these ten components would lead to improvements in teaching and learning and hence improved student performance – was not made explicit. Teams of central office staff, school staff, and community members developed plans that set forth the details for implementing the reform. Based on examination of these plans, other statements made by Superintendent Hornbeck and other district officials, and the actions taken by the district after the plan's adoption, the evaluators described the plan's theory of action as follows (CPRE et al. 1996):

If the district

- works with the schools and the community to set high academic standards for student achievement;

- aligns assessment with those standards;

- establishes an accountability system that offers strong incentives;

- delegates more authority over school resources, organization, policies, and programs to the schools;

- monitors equity throughout the organization; and

- builds public understanding and support for reform;

and if central office and the clusters

- provide guidance and high-quality support to schools and small learning communities;

then the teachers and administrators of the Philadelphia schools, in consultation with their

communities, will be motivated to develop, adopt, or adapt instructional technologies and patterns of behavior that will help all children reach the district's high standards.

The superintendent accepted this as an adequate summary of his reform ideas.

This theory of action highlights some key beliefs and values – articulated in district documents, in speeches made by Superintendent Hornbeck and other leaders, in interviews with district staff, and in discussion at policy meetings – that underlie the reform. They included:

All children can learn to high standards. The central tenet of Children Achieving was that "All children can learn, and 'all' means 'all.'" "All" included classified students, second-language learners, and all students at risk of poor performance.

The focus should be on results. To Hornbeck and his supporters, results were what mattered; how they were achieved was, at least in theory, less important.

Equity is paramount. The school district must be an advocate for the poor children it serves. Equity – of academic expectations, learning opportunities, and achievement outcomes – was a paramount objective.

School personnel need autonomy to meet the needs of their students. The theory of action and the work plans were based on an assumption that those working closest to students knew what was best for them, and wanted and needed the freedom and authority to act on this knowledge. Hence, central authorities in the district should not prescribe the means to achieve the goals lest they inhibit decisions and action by school staff.

Strong incentives are necessary. To spur action at the "cluster"[2] and school level, strong incentives had to be developed. Incentives included rewards and sanctions for performance as well as for adopting particular strategies or behaviors.

Do it all at once. Reform in all aspects of the system had to occur simultaneously and immediately to achieve significant results.

Not all of these beliefs and values were consistent, nor were they given equal weight or consistently apparent over the course of the reform. But the emphasis on being comprehensive and systemic (do it all at once) was strongly held throughout the reform and presented a challenge to formative evaluation. The district leadership's belief in comprehensive reform was so strong that, even though formal and informal opportunities for feedback to district policy-makers were frequent, comments about the confusion created in the schools by the simultaneous rollout of multiple reform activities were generally disregarded. District leaders felt that the benefits of integrated reforms would be lost if their implementation was sequenced over time.

We will return to these beliefs and values throughout this chapter, demonstrating how they shaped policy development and reform implementation; relationships among the schools, clusters, and the central office; and roles and decisions of central office leaders.

The Critical Drivers of the Reform

The critical drivers in the theory of action were standards, accountability, and decentralization. The reformers believed that these policy levers would energize the district and motivate staff and students to work towards higher performance.

Standards

Content standards were a cornerstone of Children Achieving.[3] Beginning in early 1996, teams of teachers were assembled to write standards in all subject areas. By late August 1996, draft standards for reading/English, language arts, mathematics, science, and the arts were distributed to teachers. Content standards in the social studies, health/physical education, and world languages followed soon thereafter.

Each set of content standards outlined the knowledge and skills that Philadelphia students were expected to acquire, with benchmarks, or performance standards, defined at the fourth, eighth and eleventh grades. In addition to requiring significant changes in curriculum, the standards also asked

2. Clusters were created under Children Achieving and are the district's intermediary organizational unit between the central office and the schools. There were twenty-two clusters, each organized around a comprehensive high school and the elementary and middle schools that feed into it.

3. Initially, performance and opportunity to learn standards were also envisioned, but they were never fully developed.

Philadelphia's accountability system was designed to assess schools' performance annually, and to reward progress or sanction decline every two years.

teachers to infuse "cross-cutting competencies" – skills and values in technology, multicultural competence, and communication – in all content areas.

It is important to note that Philadelphia's content standards did not specify a curriculum. Though they superseded the previous administration's "Standardized Curriculum," which prescribed a scope and sequence by grade level, the content standards simply defined broad parameters within which teachers and principals were expected to design their own curriculum. While consistent with the theory of action, this turned out to be an Achilles heel of systemic reform in Philadelphia.

Accountability

Philadelphia's accountability system, the Professional Responsibility Index (PRI), was designed to assess schools' performance annually, and to reward progress or sanction decline every two years. The PRI was made up of five indicators: student performance in reading, mathematics, and science as measured by the Stanford Achievement Test, ninth edition (SAT-9); a combined measure of teacher and student attendance; and the promotion rate (for elementary and middle schools) or the persistence rate (for secondary schools). These indicators were combined mathematically into the PRI, which provided each school with an annual score and with improvement targets.

The baseline year for the PRI was 1995–1996. Biennial targets were set for every school that assumed each school would make consistent, linear progress from its own baseline. New baselines were calculated every two years. The ultimate goal was for all schools to achieve or exceed a score of 95 on the PRI (out of a possible 120 points) by 2008. This score would mean that the average child in every school was achieving proficiency in the core subjects as measured by the district tests. Schools that met or exceeded their biennial targets were to be rewarded with cash; schools that did not meet their targets would receive assistance. If these interventions failed to bring improvement, schools faced reconstitution, the ultimate sanction in this scheme. Although two high schools were identified for reconstitution in 1997, this sanction was not employed in the first five years of Children Achieving.[4]

The accountability plan included the development of more challenging promotion standards for students at grades four and eight and new end-of-course examinations for core high school courses, but these were to be phased in beginning in 2000. New curriculum-related assessments had to be developed, and the superintendent and board of education, acting on the belief that students should not be subject to sanctions without appropriate supports, made the adoption of promotion standards contingent on securing additional funding for student supports such as an extended school day and summer school. The development of the new assessments began in 1999 and they were field-tested in the spring of 2000. However, when the deadline arrived for their adoption as promotion requirements, the district lacked the resources to provide the promised supports, and implementation was further delayed. This action kept intact the principle of linking pressure with support; however, the accountability provisions in Philadelphia ended up unbalanced for five years, falling heavily on teachers and school administrators but initially less so on the students whose effort was required to improve achievement.

4. Soon after the 1996 baseline scores on the SAT-9 were announced, the school district also announced its plans to reconstitute two high schools. The Philadelphia Federation of Teachers was outraged and charged that the reconstitution plans had been made without the appropriate consultation and before setting mutually agreed-upon criteria. An independent arbitrator agreed with the union and the reconstitution plans were abandoned, but not without cost. The episode seriously disrupted the two high schools marked for reconstitution (the principal of one of the schools had her car vandalized and was the subject of threats for her support of Hornbeck's plans) and embittered an already tense relationship with the teachers' union.

Decentralization

Along with standards and accountability, the other primary driver of the reform was decentralization. As conceived in Philadelphia, decentralization had four major components:

- *Small learning communities.* Small learning communities were intended to improve the conditions of teaching and learning in all schools, to strengthen relationships among teachers and between teachers and students, and to be the primary vehicle for improving instruction. They were subunits of schools, typically including four hundred or fewer students across several grade levels as well as the teachers responsible for their instruction. Some of Philadelphia's high schools and middle schools had voluntarily experimented with similar strategies prior to David Hornbeck's arrival, but small learning communities had not spread across the district until mandated as part of the Children Achieving reforms.

- *Local School Councils.* Each school was expected to establish a Local School Council (LSC) comprised of teachers, parents, the principal, and, at the secondary level, two students. The councils were to oversee school policies, review the budgets of small learning communities, and develop action plans to involve parents and communities in their schools to help improve student achievement.

- *Clusters.* The district was divided into twenty-two clusters. Cluster offices were to work directly with schools in support of reform. In Hornbeck's view, they were the "engines of change." Cluster offices had staffs who worked with a comprehensive neighborhood high school and the middle and elementary schools in its feeder pattern. The size of the cluster offices varied from a half dozen staff to over twenty depending on the cluster's capacity to raise external funding. The first six clusters were established during spring 1995 and the remaining sixteen in fall 1996. Clusters were expected to play a catalytic role in school improve-ment, guide and monitor the implementation of the reform agenda, provide focus for improvement initiatives, supervise principals, energize the schools, and mobilize resources. They also were expected to provide professional development, coordinate social services for schools, and strengthen K–12 articulation.

- *A streamlined central office.* The blueprint for the Children Achieving initiative clearly stated that the functions of the central office would be limited; it would "set standards, assess progress, monitor for equity, and act as a guide and provider of resources and support" (School District of Philadelphia 1995, p. iv.). This newly streamlined central office would give schools and clusters the freedom to make instructional decisions and put in place an infrastructure to help them make good decisions.

Through these changes in the structure and organization of the school district, the architects of Children Achieving hoped to increase the commitment of educators and parents and to improve productivity by reallocating power and resources and by reducing the isolation of teachers and school administrators.

Supports for Reform

District leaders recognized that Philadelphia's teachers and administrators would need new knowledge and new tools to implement these reforms and that they would need considerable support to do it well. They devised supports to help schools implement the standards and meet their performance targets. These included expanded professional development for teachers, curriculum frameworks, and family and community supports for students.

Professional Development

The Office of Leadership and Learning (OLL) was charged with developing and implementing a plan for professional development for administrators and teachers. It also was responsible for identifying and disseminating "best practices" – research-based reforms aligned with the new content standards.

The Teaching and Learning Network (TLN) was part of the OLL and served as the professional development arm of the district. However, TLN coordina-

tors and facilitators were based in the cluster offices and took their primary direction from cluster leaders, rather than from district staff. The number of TLN staff in a cluster varied from a few to over twenty as a result of differential district funding of the clusters and their varying capacity to obtain external funding. The TLN staff were expected to help schools and teachers implement district mandates and programs and to support the improvement initiatives of the schools and clusters. They offered workshops and classroom coaching for teachers. Largely recruited from the Title I program, most of the TLN staff had elementary backgrounds and lacked the content knowledge needed to help middle and high school teachers. Moreover, the pressures of implementing many changes simultaneously meant that they were usually forced to provide broad coverage rather than intense support. And the high levels of teacher turnover in some schools meant that their support often had to be focused on new teachers.

In response to concerns from teachers that the standards appeared divorced from content and pedagogy, making it difficult for them to develop curriculum, the district sponsored summer content institutes – weeklong professional development sessions in each core discipline linked closely to the standards – in the summer of 1997, a year after the standards had been adopted. They were well received by teachers, and participation increased dramatically over the course of the reform.

Districtwide Curriculum and Instructional Initiatives

Curriculum Frameworks. Developed in spring 1998 in response to teachers' requests for more guidance on how to implement the district's standards, the Curriculum Frameworks offered examples of instructional activities, units of study, and assessment tools for the standards in each subject area for every grade. They did not provide a coherent curriculum, however, nor did they provide a scope and sequence. While teachers found them useful, they continued to ask

for more specific curricular guidance and for instructional materials related to the standards. Thus, by the fifth year of Children Achieving, central office staff were beginning to develop a districtwide curriculum.

Other Instructional Initiatives. Two other districtwide initiatives provided materials and sustained professional development to Philadelphia teachers during the Children Achieving era. One was the Philadelphia Urban Systemic Initiative, a five-year (1995–2000) systemic change effort funded by the National Science Foundation (NSF) with the goal of raising mathematics and science achievement for all students. Its strategies for change included providing teachers with intensive curriculum-based professional development and standards-based materials, promoting effective programs approved by NSF, and creating and supporting a network of teacher leaders. Interviews and observations of teachers indicated that this strategy was quite effective. Teachers who participated in USI activities were more supportive of the reform and much more likely to be implementing standards-based practices in their classrooms. This might have been because they were receiving curriculum units and content-specific professional development, neither of which was available to other Philadelphia teachers in the early years of the reform.

The second initiative, Early Balanced Literacy, was undertaken by the district in 1998 to ensure that children would leave the primary grades with a strong foundation in reading and writing. In the early years of Children Achieving, a number of elementary schools adopted or developed early literacy programs using a balanced phonics/whole language approach. Based on the success of these schools, the central administration made early literacy a districtwide focus and provided participating schools with materials and professional development as well as literacy interns to reduce class size. The Annenberg Foundation also provided additional funds to support early literacy programs in several clusters.

Family and Community Supports

The Family Resource Network was led by cluster staff and included school personnel such as nurses, guidance counselors, and teachers. It sought to strengthen student support services by mobilizing and coordinating community-based agencies and

direct service providers. Together with school personnel, they were expected to provide the "safety nets" that so many poor children need. The superintendent also proposed that city and private agencies work together to ensure that all students entered school ready to learn by expanding early childhood opportunities. Although the district successfully implemented full-day kindergarten systemwide, the envisioned early childhood initiative never got off the ground.

Research as a Support for Reform

The evaluation of Children Achieving was as comprehensive as the reform itself. Asked to play the roles of both formative and summative evaluators of the initiative, researchers from the Consortium for Policy Research in Education and Research for Action were expected to track the implementation of all the major components of the reform and to document their impact on classrooms and schools and student outcomes. Between 1995 and 2001, the research team interviewed hundreds of teachers, principals, parents, students, district officials, and civic leaders; sat in on countless meetings at which plans were designed, debated, and revised; observed the implementation of reforms in classrooms and schools; conducted and reported on two systemwide surveys of teachers; and carried out independent analyses of the district's test results and other indicators of system performance.

Regular oral feedback and periodically released written reports were provided to district leaders. Indeed, one of the supports for reform was our ongoing feedback on its progress in the schools. This was particularly true in the first several years of reform, when the role of the Children Achieving Challenge (the fiscal agent for the initiative), was particularly strong. For example, the initial report on the Family Resource Network led to a reorganization of the work of the Network; broader efforts to inform teachers of the new roles that school-based student-support

professionals were supposed to play; and an emphasis on the collaboration necessary among student-support and instructional staff in schools. The district was also advised on needs that were not met by the original strategic action design, which led to the creation of summer content institutes and the development of curriculum frameworks. And problems were noted in the design that had been generating resistance or blocking implementation, such as the recruitment of TLN staff and the treatment of maternity leave as absenteeism in the accountability system.

As the Challenge's role weakened, our critical input had less impact. District leaders continued to seek input on specific issues that needed attention and were happy when the findings shed a positive light on an element or outcome of the reform. However, when the feedback challenged a strongly held belief or a core aspect of the reform, the information seemed less welcome. District leaders always listened to the findings or read them carefully, but, over time, fewer and fewer recommendations were implemented.

IMPROVEMENT AND
CHANGE OVER THE COURSE
OF THE REFORM

In this section we describe the impact of the reform program under Children Achieving on curriculum, instruction, and student performance. In the five years of Children Achieving, student test scores in Philadelphia as reported by the district rose significantly, although unevenly.[5] Gains were greatest in the first two years of the reform; they began to plateau in the third and fourth years. In the baseline administration of the SAT-9 tests in 1996, averaging across all subjects and grades, only 29.9 percent of the students tested scored at the basic level or above. The percentage scoring at this level rose to 41.9 on the 1999 tests. Table 1 presents the test results by subject and grade level. While gains were made in all subjects and at all levels, the improvement was most consistent in the elementary and K–8 schools.

The improvements displayed in Table 1 are especially noteworthy given that Philadelphia also aggres-

5. Philadelphia began using the SAT-9 in 1995–1996, the first full school year of Children Achieving. This analysis includes data from spring testing in each year from 1996 to 1999.

sively promoted the testing of all students. Compared with other urban school districts in this period, Philadelphia had one of the most inclusive testing policies, testing many special education students and English-language learners. From 1996 to 1999 the proportion of eligible students tested increased by 16 percent.[6] Since the students who were untested in the initial year of Children Achieving were likely to be lower achievers on average than those who were tested, the increased participation in the testing program undoubtedly acted as a drag on districtwide performance. Yet test scores rose significantly, in spite of the increased inclusion of these lower-performing students.[7]

However, there was one feature of this analysis that needed to be examined carefully. The major gains reported in student performance on the SAT-9 were based on an increase from year to year in the *percentage of relatively high-achieving* students (i.e., students who scored at basic or above) with respect to the *whole school,* including untested students. An analysis based only on the *average* performance of tested students (i.e., not including untested students) showed gains that were still statistically significant but not quite as dramatic. The reason for this apparent discrepancy is the influx of formerly untested

students. As a higher and higher proportion of students were tested, both the percentage of students scoring at or above basic *and* the percentage of students scoring below basic increased in many schools, especially in middle and high schools, where the proportions of untested students were largest.

To illustrate, we provide this hypothetical example: In a school's first year of evaluation, 30 percent of the students were untested; 35 percent of students scored at or above basic; and 35 percent scored below basic. In the second year, after a campaign to include more students in the testing, only 10 percent of students remained untested. Most of the 20 percent of students that were newly tested in the second year scored below basic, but improvements in the rest of the test-takers moved a net of 10 percent of students from below basic to at or above basic. Therefore, the second year's figures would be: 10 percent untested, 45 percent at or above basic, and 45 percent below basic. While it was true that the *percentage of students performing at or above basic* increased from 35 to 45 percent, it was also true that the percentage of students performing *below* basic increased, so the *average* performance did not increase or increased more modestly.

Figure 1 compares these two ways of analyzing the Philadelphia SAT-9 data. Figure 1A shows the percentage of Philadelphia students scoring at or above basic on the SAT Reading test from 1995–1996 through 1998–1999, by school level. The percentages rose significantly from year to year. Figure 1B shows the average student performance as measured by the mean normal curve equivalent (NCE). Increases are still apparent, but they are smaller.[8]

The most robust gains were made in elementary schools, followed by middle schools. The average performance of eleventh-graders was flat over the course of the reform.

6. Under Children Achieving, schools that did not test all eligible students were penalized in the district's accountability system.

7. While these improvements in achievement were encouraging, it must be noted that the overall performance of students in the district remained low relative to other Pennsylvania districts. By 1999, the average eleventh-grade reading score was still more than 150 points below the state average, and the mean fifth-grade reading scores were nearly 200 points below the state average. (Data from Pennsylvania State System of Assessment: <*www.paprofiles.org*>.)

8. Raw scores on a given test are transformed into "norm-referenced" scores, of which the NCE is an example, to establish how an individual or group scored in comparison to a "norming group." For the SAT-9 that was used in Philadelphia, the norming group is a nationally representative sample of students at a particular grade level (e.g., U.S. eighth-graders). The NCE ranges between 1 and 99 and has a standard deviation set so that at three points – 1, 50, and 99 – the NCE would be equivalent to its corresponding percentile ranking, another type of norm-referenced score. In any test results reported using the NCE, including the SAT-9, we know that the average score is 50 and that half of the test-takers in the national norming group score above the mean and half below. NCEs are useful, like other norm-referenced scores, because they meet the statistical assumption of linearity and can be used in many linear analyses, unlike the percentages also reported in this chapter.

School Level	Reading					Math					Science				
	1995–1996	1996–1997	1997–1998	1998–1999	1999–2000	1995–1996	1996–1997	1997–1998	1998–1999	1999–2000	1995–1996	1996–1997	1997–1998	1998–1999	1999–2000
Elementary	40.4	48.7	54.4	55.1	57.7	32.2	41.9	45.3	47.8	49.1	37.1	46.4	52.5	52.2	61.5
K–8	59.9	65.1	70.4	71.3	n/a	40.4	48.5	53.7	54.5	n/a	42.1	51.9	57.6	59.3	n/a
MIddle School	43.3	50.5	55.0	58.5	55.8	15.7	18.5	25.2	24.4	23.2	18.1	23.7	31.4	30.2	27.4
High School	25.6	34.0	33.6	37.0	40.7	11.5	13.8	15.9	15.2	16.4	4.9	8.0	8.4	12.0	19.1

Table 1. Percentage of Philadelphia students scoring at or above basic on the SAT-9, by subject area and school level, 1995–1996 through 1998–1999
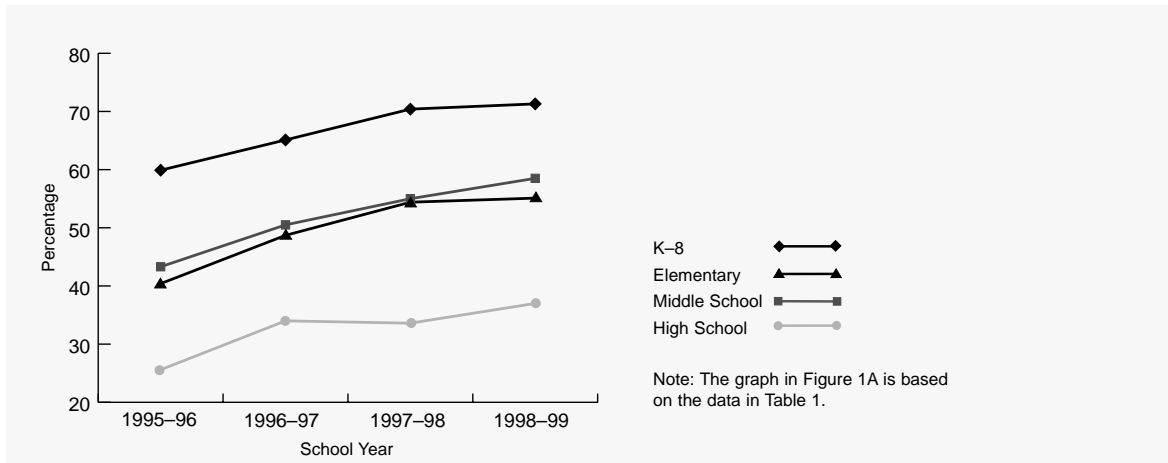

Figure 1A. Percentage of students scoring at or above basic on SAT-9 reading

	1995–1996	1996–1997	1997–1998	1998–1999
4th Grade	40.1	42.0	44.7	43.9
8th Grade	43.9	43.1	45.3	46.9
11th Grade	36.9	36.2	35.2	36.6

Note: normal curve equivalent scores were not available to us for the year 1999–2000. This figure shows data through the 1998–1999 school year.
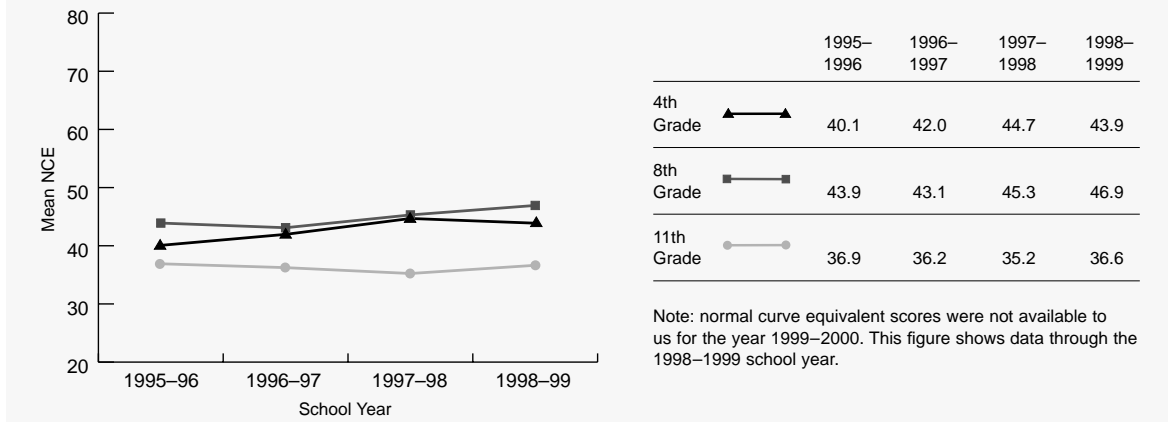
Figure 1B. Average SAT-9 reading scores as measured by the mean NCE

Figure 1. Varying interpretations of improvement

Attendance, Promotion, and Persistence

Statistically significant gains were also made on all of the other indicators of school performance. The percentage of students in attendance for 90 percent or more of school days increased by three percentage points from 1996 to 1999; and the percentage of staff attending 95 percent or more of school days increased by over 6 percentage points. Persistence (on-time graduation) and promotion rates also increased significantly but less rapidly.

Relationship of Achievement Gains to Reforms

How can we account for these changes in performance? To what degree are they related to the implementation of the reforms? What aspects of the Children Achieving reforms seem to account for the achievement gains? Were the gains largely a response to the high-stakes accountability system? This section uses qualitative and quantitative evidence to address these questions.

As we have seen, Philadelphia elementary students made the most consistent gains on the SAT-9. Our quantitative[9] and qualitative research suggests

that three factors contributed to these gains in elementary schools:

- test preparation
- focus on literacy programs in the primary grades
- development of strong professional communities in the schools

Test Preparation

Our qualitative data indicate that, in all likelihood, improvement in student achievement at all levels was to a significant degree the result of intensive test preparation and teachers' increasing familiarity with the content and format of the test.[10] While teachers initially saw inadequate curriculum guidance as a serious problem, the preeminent role of the SAT-9 test quickly shifted their attention from the content standards to the content of the test. Various forms of test prep were observed at all levels and were the most common instructional response to the reforms. Not all of this test prep was bad. Some of what might be considered "test preparation" were in fact educationally sound improvements in the curriculum to reflect the expectations in the test. These included an increase in writing assignments and higher standards for student writing. However, some of the test prep observed in Philadelphia was of the "drill and kill" variety. Teachers used materials such as Harcourt-Brace's KeyLinks workbooks to develop students' test-taking skills and familiarize students with the test.

Furthermore, the same form of the SAT-9 was used each year. While the district made an extraordinary effort to keep the test secure, researchers found copies of the test in the schools, and some teachers were familiar with the open-response questions. This problem arose again when the district field-tested its new fourth- and eighth-grade promotion tests and end-of-course exams. Copies of these new exams were readily available in the schools after they were administered, and the same tests were used in the subsequent year. We cannot say what effect the test security problems had on performance, but it is logical to assume that there would be some effect.

Focus on Literacy Programs in the Primary Grades

While at least some of the achievement gains can be attributed to the test preparation activities, the evidence suggests that instructional improvements also

9. Using hierarchical linear modeling (HLM), we examined the relationships between teacher and school characteristics, measures of professional community, measures of reform implementation, and growth in test scores over four years (1995–1996 through 1998–1999). The analysis showed that both the poverty level of students and the degree of professional community in a school were directly related to its growth in achievement in fourth grade. Interestingly, poverty did not depress growth in achievement. Schools with the highest concentration of poor students actually improved faster than schools with lower concentrations of poverty. But poverty was also a significant predictor of the baseline scores, so these poorest schools also had lower baselines and therefore more room for improvement. The only measure of professional community that was significantly related to growth in achievement was teacher collaboration: schools with greater teacher collaboration experienced higher rates of growth in achievement. The measures of reform implementation were not significant predictors of achievement growth in our HLM models, but subsequent analyses using logistic regression did reveal significant relationships between small learning community implementation and school conditions, as well as between small learning community implementation and professional community measures. For a more detailed explanation of this analysis, please contact the authors.

10. Efforts to create a quantitative measure of the level of test preparation in schools were not fruitful.

played a role in the gains and in their unevenness. In particular, qualitative data indicate that Philadelphia's focus on early literacy paid off in the primary grades. Classroom observations showed that teachers in the early grades increasingly used a balanced approach to teaching reading and writing, cooperative groups, and an emphasis on drafting and revising.

In contrast, Children Achieving did not offer middle and high school faculties equally specific or effective approaches to instruction. For the most part, middle and high school teachers did not focus on one or two robust and substantive strategies for improvement in student achievement. They faced large numbers of students who lacked basic skills and who required considerable support and remediation. Many students were alienated and hard to engage in academic work. In response to district mandates, teachers tinkered with the structural arrangements, creating small learning communities, interdisciplinary curricula, project-based learning, and service learning. In middle schools, they created new curriculum tied to their small learning community themes. After five years, we judged most of this thematic curriculum work to be still at an early stage of development. It seldom involved students in rich intellectual work, nor was it informed by multidisciplinary perspectives. High school faculties expanded opportunities for students to participate in internships and service learning, but were less successful at making classrooms more challenging learning environments or stimulating deep changes in instructional practice.

Strong Professional Community

While our analysis found no direct relationship between the degree of implementation of the Children Achieving reforms and growth in student achievement, we did find that the relationship may have been indirect. We found that well-implemented small learning communities were connected with teacher reports of higher levels of professional community,[11] and that there was a relationship between the strength of the professional communities, positive school conditions, and improved student achievement in elementary schools. That is, our analysis suggests that the implementation of small learning communities was associated with higher levels of professional community and that higher levels of professional

community were linked to improved student achievement (controlling for significant factors such as poverty). Given the limitations of our data, no causal relationship can be inferred, but the findings do suggest possible directions for future work in the district and for further research.

The data also suggest that, in some schools, strong professional communities and positive school climates preceded Children Achieving and offered fertile ground for the creation of small learning communities and substantive pedagogical change. This was the case in at least two of the elementary schools and one middle school of the twenty-one schools[12] where we conducted intensive, multiyear qualitative fieldwork.

While small learning communities appeared to provide some benefits, they also generated new problems. In schools in which teachers were permitted to choose which small learning community they wanted to join, inequitable distributions of teaching talent sometimes resulted. The hardest-working, most able teachers often chose to work together, leaving some small learning communities staffed by those more resistant to the reform or less motivated. Students in most high schools were able to rank the small learning communities by their "quality," evidence that a form of de facto tracking had emerged. In addition, pressure for students to take most of their classes within their small learning communities meant that students' access to the full curriculum varied. This was most obvious in science, mathematics, and foreign languages, where not enough specialists were available to provide equivalent staffing in all small learning communities. A related problem was the variation in the quality of curriculum and curriculum implementation across small learning communities that followed the demise of middle and high school academic departments. Content teachers were in many cases on their own, with little access to support from peers in their fields.

11. We measured "teacher professional community" through our teacher survey. It encompasses items that describe the dynamics among teachers, teacher-principal relations, teacher collaboration, and principal leadership (see Tigue, Wang & Foley 2002).

12. This sample of twenty-one schools included eleven elementary schools, five middle schools, and five high schools.

LESSONS LEARNED FROM CHILDREN ACHIEVING

Children Achieving was a significant reform effort in both the amount of political capital expended and the investment made, but its effects were modest at best, and it was not sustained. What lessons can be drawn from this experience? Here we offer the insights we have gained from our five years of studying the reform in Philadelphia. They are not new insights, but they bear repeating because of their importance to the success of school reform in any school district, and to the role evaluators might play in the implementation of a reform strategy.

The Importance of Context

Philadelphia was not a wealthy city in the 1990s. Its population had decreased dramatically from the 1970s to the 1980s, and so had its middle class tax base. During that decade, the total population of the five largest U.S. cities – Philadelphia among them – decreased by 9 percent, while the population living in poverty grew by 22 percent (Wilson 1987, p. 46). When David Hornbeck began his tenure as superintendent, the city was still recovering from a serious economic crisis.

With that history, the city refused to provide significant additional resources for Children Achieving, arguing that it had "stretched its taxing ability to the limit" (School District of Philadelphia 1998, p. 26). But the full implementation of Children Achieving required significant additional funding, more than the $30 million generated annually by the grant,[13] and its design assumed that more funding would be forthcoming. In launching the initiative, Superin-

tendent Hornbeck and his supporters took a calculated risk that the Annenberg Challenge grant would enable them to improve the performance of the system, and that evidence of improved performance would generate the political will needed to obtain increased funding through the city, the courts, or the legislature.

By 1997, the superintendent, the board of education, the city council, and the mayor all agreed that the Commonwealth of Pennsylvania was not upholding its fair share of the costs of educating Philadelphia's students. But state officials did not see it that way.[14] They believed that funds were being used inefficiently in Philadelphia and that the district's teacher contract was a major obstacle to improvement. In their view, better management and a better contract were prerequisites for additional state funds. The school district and the city used many strategies to induce the state to provide additional funding – multiple lawsuits, political brinksmanship, public scolding – but to no avail. The annual fiscal crisis of the school district became one of the few constants of the Children Achieving reform era.

Without new financing from the state, per pupil funding in Philadelphia remained well below what was spent in the surrounding areas. In 1997, Philadelphia spent $6,812 on each public school child. When compared to wealthy suburban school districts such as Jenkintown, Lower Merion, and Radnor, the gap was as much as $5,443 per student (School District of Philadelphia 1998, p. 11). Teacher salaries were also higher in suburban areas. Starting salaries in the suburbs were more than $3,500 higher than starting salaries in Philadelphia and maximum salaries were more than $9,000 higher (ibid., p. 29). Average teacher salaries in Philadelphia also fell below statewide teacher salary averages.

According to the school district, expenditures on administration declined during the Children Achieving era, although to make that claim they had to count the cost of the Teaching and Learning Network (TLN), professional development specialists based in the cluster offices, as an instructional expense. Critics who saw the TLN staff as an "administrative" cost contended that expenditures on administration actually grew over the course of Children Achieving.[15]

---

13. Though a significant source of discretionary funds, the $30 million from the Annenberg Challenge grant was equal to only about 2 percent of the $1.5 billion annual budget.

14. The funding that Pennsylvania provides to each school district currently is based on a funding formula which takes into account the number of pupils, the special needs of the district, its ability to raise local taxes, and other factors. However, the state froze the formula in 1993, which meant that state aid to the district did not rise in response to increases in enrollment and poverty in Philadelphia. In per pupil dollars adjusted for inflation, the real value of state education funds coming to Philadelphia annually between 1993 and 1998 actually decreased by 5.9 percent (see Century 1998).

15. See, for example Snyder 1998; Kirsch 1998.

There is no doubt that the number of staff assigned to the central office was smaller at the end of the reform than it was before it, but with over two hundred staff assigned to the cluster offices, school personnel felt that there was more bureaucracy, not less.

Inadequate resources limited the school district's ability to provide time for teachers and other personnel to receive professional development, to develop curriculum, and to work with colleagues. They also hampered the district's ability to hire the most qualified personnel. Teachers, in particular, had to make a real commitment to urban education (or be unable to obtain a job in the suburbs) to accept the lower starting salary in Philadelphia, a salary that was further reduced by the city tax on wages. Scarce resources also limited the ability of the school district to provide up-to-date curriculum materials and technology.

The Need to Build Constituencies and Partnerships

Social capital is a product of relationships among people. For example, a group of people who trust each other has a form of social capital. All other things being equal, a trusting group is more likely to succeed at a given task than a group whose members do not trust each other (Spillane & Thompson 1997). In Philadelphia, the limited social capital in the school district and inadequate efforts to build stronger constituencies for reform affected the implementation of Children Achieving. In particular, the culture of the district, its history of reliance on line authority, and the relationship of the central administration with its potential partners, including the Philadelphia Federation of Teachers, its own cluster leaders, state officials, and the business community, all affected the supply of social capital.

The Philadelphia Federation of Teachers. The school district's relationship with the Philadelphia Federation of Teachers (PFT) suffered over the course of Children Achieving and was characterized by mistrust on both sides. In a sense, the school district administration was at war with its own teachers. The PFT strongly objected to key components of the reform plan, particularly to its accountability provisions. They objected to spending money on cluster staff when schools were understaffed. The leaders of the PFT felt that Children Achieving was a threat to the union and to hard-won work rules in the teacher contract. Tensions were highest when the school district administration attempted to reconstitute two high schools; as a result of union objections, plans for the reconstitution were ultimately halted by an independent arbitrator who ruled that the district had failed to engage in the necessary consultation with the teachers' union. To the PFT leadership, the reconstitution attempt was just one example of the Hornbeck administration's pattern of excluding them from the decision-making process.

School district leaders, for their part, told us that the PFT representatives were invited to meetings about relevant policy areas, but that they either obstructed the meetings they attended or never showed up. Central office leaders felt that the PFT leadership was adversarial and unreasonably attached to the unproductive rules and regulations of an antiquated contract, and that the PFT had the interests of teachers, not children, at heart. In our estimation, both groups shared the blame. In four years of meeting with and interviewing central office staff and PFT representatives, we seldom heard positive comments from members of either group about the other and frequently encountered distorted interpretations of the other party's motives.

The acrimony on both sides of this relationship made progress difficult. The school district and the PFT were unable to agree on contractual changes that would have supported Children Achieving, especially in the area of decentralization. School communities were not permitted to select their own principals and staff, as Children Achieving advocated, and there was conflicting language about local school councils in the Children Achieving plan and the PFT contract.[16] Additionally, the failure of the school district to gain

16. The language of Children Achieving called for 35 percent of households to vote to determine council membership; in the teachers' contract the provision was for five parents to be selected by the Home and School Association. Additionally, Children Achieving called for two-year terms for parents, while the contract outlined one-year terms for teachers. See Christman 1998.

concessions from the PFT undermined its credibility with a number of stakeholders, particularly principals and the business community.

The tension between the PFT and district leaders also made the job of evaluation more difficult. PFT leaders viewed our frequent meetings with district officials with suspicion. Though we were providing formative feedback to the district, often mentioning the need to develop better relations with the Teachers' Union, the PFT questioned our impartiality. In turn, our willingness to hear out the PFT's questions and objections to the reform, and to seek their cooperation in administering surveys of teachers, made some district leaders apprehensive, and may have reduced our access to them in later years of the reform.

Cluster Leaders. In addition to an antagonistic relationship with the leadership of the Philadelphia Federation of Teachers, the school district often alienated its own cluster leaders, who were crucial to reform implementation. Cluster leaders were, for lack of a better term, regional superintendents, who were supposed to improve and align instruction across a feeder pattern of schools and lead and support local professional development and community engagement. They were members of the superintendent's cabinet, which also included key central office leaders.

With the addition of twenty-two cluster leaders, the cabinet ended up being a group of about fifty people, a size that was ill-suited for collaborative work. Cluster leaders came to describe cabinet meetings as the place they came to talk about decisions that were already made by central office staff. Cabinet meetings were also one of the few forums they had to air their grievances and, as a consequence, central office staff often felt "ganged up on" by cluster leaders.[17] In one particularly contentious meeting, for example, cluster leaders were upset that more information was not available as to how they would finance and organize summer school programs, scheduled to begin about three months from the time of the meeting. They made little effort to hide their anger and hostility.[18]

This tension arose in part because of conflicting ideas about the cluster role. Whereas some central office staff saw clusters primarily as vehicles for informing the field about new aspects of the reform, cluster leaders felt they should have the autonomy to determine the means of improving performance. In interviews, they frequently complained about central office mandates and their lack of influence over policy. In the following excerpt from our field notes, a cluster leader illustrates this point:

> Part of the challenge I have had as a cluster leader is to keep the central office away from me, so I can allow my people to develop their responsibilities. Downtown keeps adding more [stuff] to our plate ....Let me give you an example. The central office wanted to change the special ed formula, which apparently they had been working on for months, but it wasn't shared with anyone [in the field]. When it was finally announced, parents went to the board and begged them not to let it happen. So the board then asked the school district what facts they have to support the change, so now *we* [cluster staff] have to do a lengthy survey. We have to identify one special ed kid per special ed classroom and review their [education plan], observe their classroom, interview the parent and teacher, and we have to do it all in four weeks. That's 75 kids for me because we have 75 special ed classrooms. The central office knew that they would ask us to do this in the summer, but they didn't actually ask us until a couple of weeks ago. This says to me that I have to put a hold on everything else I'm doing and do this. It takes away from your focus.[19]

The central office staff became aware of time and turf concerns and made efforts to seek cluster leader input and plan with respect to cluster schedules. Nevertheless, the unproductive relationship endured, and cost the district key support. In the 1999–2000 school year, the superintendent asked the cluster leaders to back him in a fight for funding from the state; they refused.

17. Personal communication, December 2000.

18. Field notes, March 24, 1999.

19. Field notes, April 1996.

State Officials and Business Leaders. The school district's relationship with state education officials, the governor, and the state legislature also was strained over the course of the reform. When Hornbeck became superintendent in 1994, there was a Democratic governor and Democratic majorities in both houses of the state legislature. He came to his position with strong backing from both Philadelphia's mayor and its business community. However, just three months into his administration, the political landscape in Pennsylvania and Philadelphia changed dramatically: the state elected a Republican governor, Tom Ridge, and Republican majorities in the state legislature who were committed to reducing government spending. Relationships between the state officials and the district were tested by the new governor's advocacy of vouchers, his refusal to grant the school district significant additional funds, and the superintendent's inflammatory rhetoric, alleging racism on the part of state officials in speeches and via a federal civil rights lawsuit against the state. When we interviewed state education department officials in the fall of 1999, their anger toward David Hornbeck was evident.

This antagonistic relationship between the state and the school district had effects on local constituencies as well. The strong backing of the business community for Children Achieving deteriorated as Hornbeck's battles with the state became more public. In addition, some of the superintendent's strongest supporters left Philadelphia as the major corporations headquartered in Philadelphia moved out. In civic organizations like Greater Philadelphia First, leadership shifted from executives of large national corporations to leaders of smaller, more local firms. The clearest sign of the fracture in the alliance between the business community and the school district was when Greater Philadelphia First – a coalition of Philadelphia business executives that served as the fiscal agent for the Annenberg Challenge – supported Governor Ridge's plan for school vouchers.

External Reform Support Organizations. From the beginning of Children Achieving, the school district had two primary external partners: The Children Achieving Challenge (CAC) and the Philadelphia Education Fund (PEF). Leaders from both groups were members of the executive committee and the superintendent's cabinet and played significant roles in the reform effort, especially in establishing and leading the seven work teams charged with developing goals, setting priorities, and creating annual work plans to implement Children Achieving.

The work teams served an important function by providing learning opportunities for the central office and cluster staff, one of the few such systematic opportunities they had. (We were part of a work team on the evaluation, which gave us many opportunities to discuss preliminary findings and get feedback on our research design.) At the outset of Children Achieving, systemic reform was a fairly new concept nationally and few anticipated the demands it would place on teachers and schools, let alone on the central office staff. In autumn 1996, for example, some central office leaders were still questioning whether standards were curriculum.[20] Most members of the superintendent's inner circle, particularly the PEF and CAC leaders, were knowledgeable and thoughtful about standards-based, systemic reform, but it took some staff at the central office a considerable amount of time to develop a deep understanding of the initiative they were helping to lead. The work teams were key to their understanding and led to important decisions about the role of the central office, the supports it would provide for the reform, and key contributions from the two external groups.

Unfortunately, the basis of the school district's partnerships with CAC and PEF were more personal than institutional. When the leaders of these organizations left Philadelphia to pursue other career opportunities, the organizations' relationships with the school district weakened. Collaboration continued, but the relationships were more marginal and less catalytic. The two leaders' departure further isolated the school district from key constituencies.

20. Field notes, November 12, 1996.

Importance of Focus

Another factor that made reform implementation difficult was the pressure on school staffs generated by the core belief that the whole system must be reformed simultaneously and immediately. The superintendent adamantly opposed piecemeal, incremental reform. He felt that the ten components of his reform plan were mutually supportive and had to be moved forward simultaneously. But "doing it all at once" created reform overload throughout the school district, from schools to the central office. School staffs were unable to focus their efforts around clearly defined and manageable instructional priorities. Cluster staffs were overwhelmed; they worked hard to win teachers' support and to assist them, but they were hampered by the sheer number of district initiatives and directives. Many clusters were unable to fully develop or implement their own reform strategies because so much time was spent promoting and disseminating information about new central office policies and programs that the schools were required to implement.

The volume of reform initiatives also overwhelmed many central office staffers, particularly those who were not among the superintendent's close advisers. The concern about overload was evident very early in the Children Achieving initiative. In an early policy meeting, when several of the superintendent's inner circle had left the room, one central office leader said, "We need to talk about priorities and make some tough, hurtful choices and let the chips fall where they may. We can't pretend any more that we can do it all." Another central office leader agreed, using the analogy: "We can't plow all the streets. Which ones are most important?" He suggested that focus should be placed on a group of schools or a few clusters.[21] But when the inner circle members returned and the other participants briefed them on what had happened in their absence, there was no mention of the concern about reform overload.

This reluctance to tell the superintendent and his closest staff about the difficulty of "doing it all at once" continued throughout the reform effort. In an interview two years after the exchange quoted above, a district leader told us:

> I've got to tell you something else. We are on innovation overload! As hard as it is for a superintendent in a large district, someone has to have the guts to say it….Everyone is tired….[Central office personnel] are having to learn something new all the time, we're rolling out so many competing forces. [Begins counting on his fingers.] We have the CSP. We have SLC. We have School-to-Career. We have service learning. We have multidisciplinary projects. And there is more to come. That is just one hand! We have judgments against us in federal courts that push us to make things not fall through the cracks….There's always a new priority.[22]

The urgency of "doing it all at once" created pressure on central office staff simply to "roll out" the reforms and move on to the next priority. There was little time to support or guide the reforms or to receive our feedback and review and revise policy. It is not surprising that, to schools and clusters, central office policy felt like unsupported mandates. The core value of "doing it all at once" increased the top-down mandates by the central office, conflicting with the core value of school autonomy.

As evaluators, we frequently tried to point out problems of this kind, but we were in an awkward position. There were many constraints on what we were able to do as witnesses to this lack of focus and still maintain the confidentiality of our informants and access to the system. Additionally, the lag time between the discovery of patterns of response such as the field's perceptions of the lack of focus – rather than merely individual instances – and their actual occurrences often made these findings less resonant with district leaders. By the time we shared our findings with them, they were often enveloped in the implementation of different aspects of the reform, a fiscal crisis, or another political battle. Moreover, the tension between the district and the teachers' union also made some of our information suspect to some in the administration. Because we had collaborated with the union on districtwide surveys, some believed that our repeated warnings about the effects of reform

21. Field notes, November 1996.
22. Field notes, December 1998.

burden, problems in sequencing, and the inadequacy of the support structures made us "apologists" for resistant teachers.

Importance of Reform Sequencing

One of the primary flaws in the implementation of Children Achieving was the sequence in which the district rolled out the reforms and supports. In order to capitalize on the momentum built up from the hiring of the new superintendent and the acquisition of the Annenberg funds and to fulfill the underlying belief in the need to "do it all at once," there was a rush to implementation. The district led with the pieces that were easiest to put in place. And, for strategic reasons, the district wanted to be able to demonstrate relatively soon that it was making gains in student achievement, to persuade the state, or the courts, that Philadelphia should receive increased funding so the reforms could be sustained. As a result, the new cluster organization, the new tests, and the accountability system were the first components to be implemented. The standards came in the next school year. The tools and supports needed by teachers to use the standards and prepare students for the tests came even later.

But to many, instituting accountability policies and other structural reforms before developing the infrastructure needed to support the changes in practice and services required to raise achievement was putting the cart in front of the horse. School personnel complained that they were being held accountable for performance targets before they had received the new standards, before all twenty-two clusters were in place, and long before the development of the Curriculum Frameworks offered a modicum of guidance and summer institutes offered teachers rich opportunities to examine their practice. All of these sequencing problems contributed to perceptions by teachers and principals that they were being asked to carry disproportionate amounts of the burden for improvement. They felt victimized by the ways in which the reforms were presented and rolled out. Faced with dire consequences, many teachers turned to sure and safe methods of instruction – drill and practice.

The urgency of "doing it all at once" created pressure on central office staff simply to "roll out" the reforms and move on to the next priority. There was little time to support or guide the reforms or to receive our feedback and review and revise policy.

Our evidence suggests that the teachers' concerns may have been valid. When the district finally put in place the instructional supports teachers had clamored for, those who were able to take advantage of them benefited from them and many improved their practice. A different sequence might have produced different results for the district.

Policy-Makers as Learners

The capacity of the central administration to support the Children Achieving reforms was an issue not only of financial resources, but also of human capital. With the exception of a few key leaders, knowledge about the substance of the reforms and how they fit together, and the expertise to implement them was limited, even in the central office. While many central office staff members were passionately committed to Children Achieving, some had only a superficial understanding of the reforms they were supposed to help schools implement and of the demands they made on teachers and school administrators. Much of what they were trying to implement existed only in theory prior to Children Achieving.

Poor personnel decisions and turnover in staff also limited central office capacity. The associate superintendent in charge of the initial development of the reform and the leader of the superintendent's transition team resigned in protest over the superintendent's insistence on promoting teacher accountability. A well-regarded central office leader was demoted for refusing to submit a resignation letter early in the

Embattled leaders attempting to implement an ambitious plan with key elements underdeveloped are not in a good position to make midcourse corrections on the basis of research information.

reform. A deputy superintendent retired. Over the course of the reform, there were three different leaders of the Office of Leadership and Learning, four directors of information technology (including two acting directors), three directors of the Office of Best Practices, three managing directors, and two directors of the Office of Curriculum Support.

Staffing turnover and ineffective leadership plagued the departments most directly responsible for providing support to the field – the Office of Leadership and Learning, the Office of Curriculum Support, and the Office of Best Practices. Staff hired to fill these vacancies were not, in general, compatible with other members of the leadership group and some gained reputations as "stallers" – people who put up obstacles to reform.[23] One central office leader admitted, "Central office personnel decisions have not been good ones."[24]

Researchers as Learners

Despite the district's stated commitment to evaluation, the experience in Philadelphia suggests that there are limits to the influence researchers can wield. Reformers working in highly politicized environments, in which the stakes are high and opponents are ready to take advantage of each mistake or sign of weakness, may be reluctant to admit to flaws in design or errors in strategy. And even if they are willing to make changes, they may be unable to do so.

23. Field notes, July 3, 1997.
24. Field notes, November 1999.

Embattled leaders attempting to implement an ambitious plan with key elements underdeveloped are not in a good position to make midcourse corrections on the basis of research information. They are committed; it is more expedient to blame the critics or the teachers who are struggling to implement their ideas than to admit error or come up with new plans on the fly.

To be sure, Philadelphia leaders did at times show a willingness to listen to criticism and a flexibility to adjust some of their plans. The district implemented well-regarded content institutes and then curriculum frameworks after teachers complained about a lack of curricular guidance. And the district delayed implementing its student promotion plan when funds for supporting services did not materialize. But for the most part, district leaders seemed unwilling to hear challenges to the fundamental tenets of Children Achieving, let alone act to change them. The response seemed to be to circle the wagons rather than build a new camp.

The lessons for evaluators seem clear. First, the same team of researchers cannot do both internal formative and public summative evaluation work in atmospheres as charged with conflict as Philadelphia was during the late 1990s. Providing formative feedback requires a high level of trust and a close working relationship with the leaders and designers of the reform. This cannot be established and maintained if the research team is also reporting to funders and sharing findings with potential critics of the reform. The high level of conflict generates a bunker mentality in which the norms are believing in the reform in toto and questioning core beliefs is tantamount to joining the opposition. Simply put, the people to whom we were charged with providing feedback were not in a position to accept constructive criticism from evaluators who also were asked to make summative judgments about the success of the reform.

Our limited impact was exacerbated by a timing problem. There was tension between the funders' desire for a broad-based view of all aspects of the reform and the researchers' interest in looking at the reform in depth. Therefore, the results from the research tended to arrive too late to be helpful to the reformers. They were either already deeply committed to a course of action or the circumstances had

changed by the time the research team reported. In retrospect, the researchers should have conducted a series of shorter-term studies focused on points of potential conflict or tension within the system that could have provided rapid and useful feedback to the designers. This would have made it harder to study the "big picture" over time, but might have been a more useful contribution to the successful implementation of the reform.

CONCLUSIONS: LOOKING BACK

By the spring of 2002, Philadelphia public education was under a new regime. The state had exercised its authority to take over the financially troubled district, and a new five-member commission appointed by the governor and mayor had assumed control. They were moving forward with plans to contract with private firms and organizations to take over some of the city's low-performing schools. The fate of the other schools was not clear. Some of the central components of Children Achieving, such as the cluster system, the PRI, and the TLN, were gone. The fate of others – the standards, the curriculum frameworks, the small learning communities, and the Family Resource Network – was not clear.

City leaders seemed to be abandoning Children Achieving and its means for bringing about improvement. Advocates of the reforms argued that considerable progress had been made and claimed that inequities in state aid, resistance from the teachers' union, and declining support from the business community had undermined the possibility of even more progress. Critics of Children Achieving in turn pointed to fiscal deficits, increased expenditures on administration, the flattening of test scores, and the emphasis on test preparation as evidence that the reforms were seriously flawed. Because so many of the critics seem to be willing to dismiss the issue of inadequate funding, it is tempting to say that the truth lay somewhere in between. In our view, though, the critics seemed to have the more convincing arguments.

Without a doubt Children Achieving had offered a compelling set of ideas for school reform and had changed the nature of the debate over public education in Philadelphia. Central ideas such as the beliefs that results matter; that all children can learn at high levels, and that "all" means "all"; that everyone must be held accountable; and that professional development is a necessity generated a new set of expectations for local policy-makers. However, the leadership of the school district of Philadelphia paid too little attention to implementation lessons from the past when they crafted the reforms. They too often criticized teachers rather than attempting to win their support. They adhered to the dictum of the Children Achieving plan that everything had to be done simultaneously, which placed enormous burdens on teachers and principals. They assumed that teachers would embrace the reforms in exchange for more freedom to develop curriculum and more influence over school decisions. They were careless in the manner in which the reforms were sequenced. They put pressure on teachers before they provided supports, and they underestimated the difficulty of developing standards-based curriculum and instruction.

Philadelphia's policy-makers also lacked a clear vision of what was required to implement standards-based instruction at the central office, cluster, and school (administrative, teacher, and parent) levels. Capacity was lacking at all levels of the system, yet efforts to build it were sporadic and weak. Left without the necessary supports and feeling overwhelmed and overburdened, many teachers, principals, and administrators left the district seeking higher salaries and better working conditions outside the city, making implementation of the reforms even more difficult.

Individual schools also varied in their capacity for change, their professional cultures, and their reform histories. The experience in Philadelphia suggests that different reform strategies are needed for elementary, middle, and high schools. Each level of school brought different organizational issues, professional norms, and cultures to be addressed. Their past experiences with reform varied, as did the challenges they faced in motivating students and staff. However,

Children Achieving only offered a "one size fits all" reform strategy that was difficult to adapt to varying school contexts.

Ideally, researchers could have brought these flaws to light and worked with the district to adapt its strategies. Problems of design and implementation are precisely those that dispassionate observers, armed with knowledge about the experiences of other districts, could best help reformers address. But the researchers' influence was limited, and the midcourse changes the district made were not sufficient to fulfill its compelling rhetoric and promises of a new day in urban education.

Yet it would be inaccurate to say that nothing was accomplished in the five years of Children Achieving. First of all, there were some real gains, such as an improvement in elementary student performance and the introduction of full-day kindergarten and early literacy programs, although the gains were modest compared to the ambition and scope of the reform effort. Secondly, the reform effort raised expectations for Philadelphia's children; it forced citizens to recognize the sobering realities of public education in Philadelphia and to debate its future.

Finally, the Philadelphia reform experience presents us, as researchers, with the opportunity to look back at the difficult lessons about what brought Children Achieving to an end and to gain valuable insights that can help future reform efforts succeed. We learned much about the challenge of sustaining systemic reform and about the need for resources and support to match the complexity and ambition of the reform design. We also learned much about the issues surrounding the role of external evaluators in a reform effort. To the extent that educational stakeholders – students, teachers, school and district administrators, the community, and researchers – can help each other learn from constructive feedback, negative as well as positive, future reform efforts will be able to overcome the challenges of the past and help public schools fulfill their high hopes and ambitious plans to educate all students to high standards.

References

Century, J. A. 1998. *Citizen's Guide to the Philadelphia School Budget.* Philadelphia: Greater Philadelphia First.

Christman, Jolley Bruce. 1998. *Guidance for School Improvement in a Decentralizing System.* Philadelphia: Consortium for Policy Research in Education.

Consortium for Policy Research in Education, Research for Action, and OMG Center for Collaborative Learning. 1996. *A First-Year Evaluation Report of Children Achieving: Philadelphia's Education Reform.* Philadelphia: Greater Philadelphia First.

Kirsch, T. 1998. "President's Message: We Are Fighting Back – Together – and We've Got to Win!" *PFT Reporter.* Available at <*http://www.pft.org/pres798.html*>.

Pennsylvania State System of Assessment. Available at <*www.paprofiles.org*>.

Philadelphia Inquirer. 1994. "A District in Distress." Special Section, G1–G8 (October 23).

School District of Philadelphia. 1995. *Children Achieving Strategic Action Design.* Philadelphia: School District of Philadelphia.

School District of Philadelphia. 1998. *Realities Converge, Revisited: School District Sees Gains on Test Scores and Management Efficiencies, But Fiscal Crisis Is at Hand.* Philadelphia: School District of Philadelphia.

Snyder, S. 1998. "No Cash Crisis in Schools: Philadelphia's Problem Is Management, Says an Accounting Firm Hired by the State Legislature," *Philadelphia Inquirer* (November 13).

Spillane, James P., and Charles L. Thompson. 1997. "Reconstructing Conceptions of Local Capacity: The Local Education Agency's Capacity for Ambitious Instructional Reform," *Educational Evaluation and Policy Analysis* 19:2, 185–203.

Tigue, Erin, Aubrey Wang, and Ellen Foley. 2002. *An Analysis of Children Achieving on Student Achievement in Philadelphia Elementary Schools.* Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Wilson, W. J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago: University of Chicago Press.