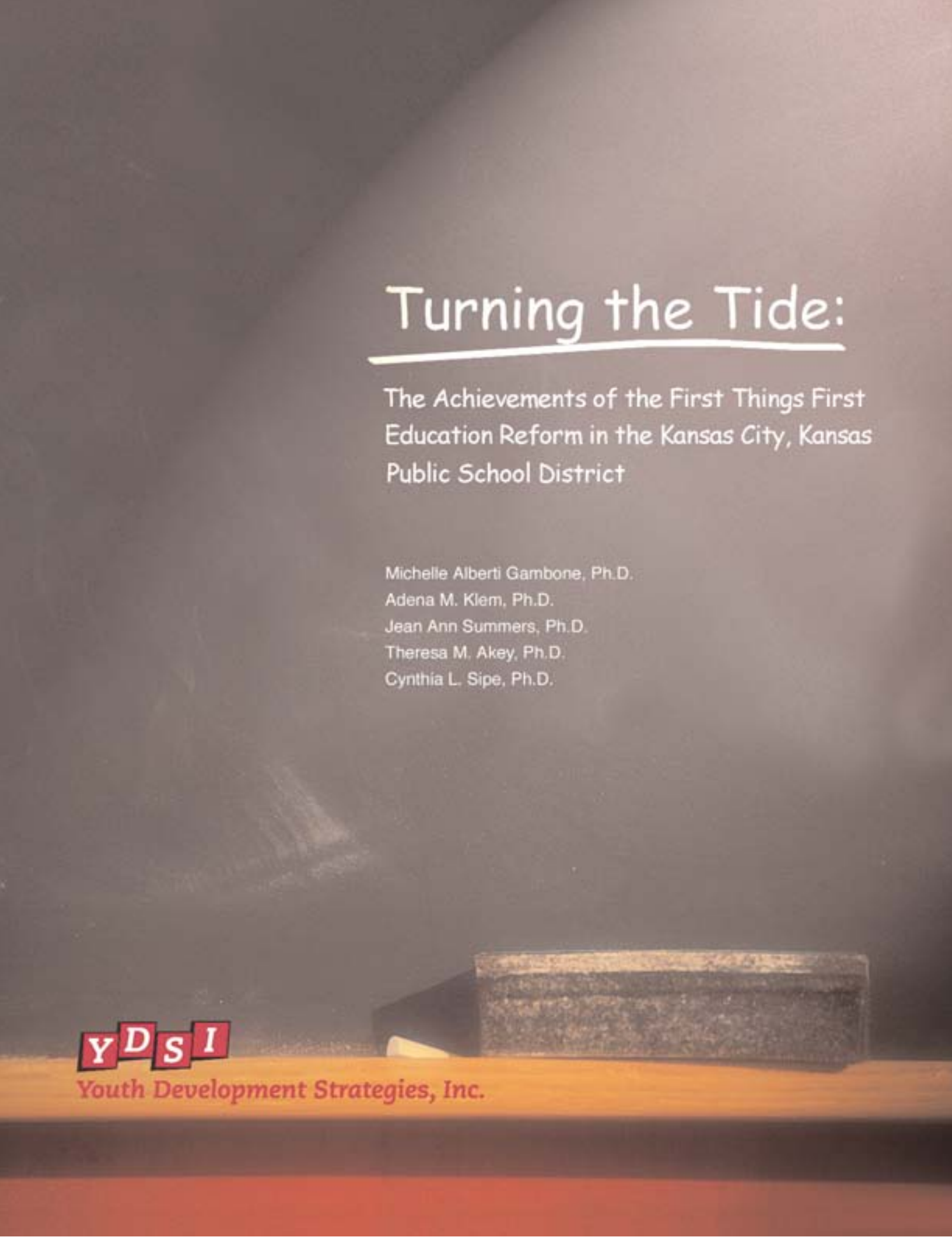


Turning the Tide:

The Achievements of the First Things First Education Reform in the Kansas City, Kansas Public School District

Michelle Alberti Gambone, Ph.D.
Adena M. Klem, Ph.D.
Jean Ann Summers, Ph.D.
Theresa M. Akey, Ph.D.
Cynthia L. Sipe, Ph.D.

YDSI
Youth Development Strategies, Inc.

This report was prepared by Michelle Alberti Gambone, Ph.D., Adena M. Klem, Ph.D., Jean Ann Summers, Ph.D., Theresa M. Akey, Ph.D., and Cynthia L. Sipe, Ph.D. for Youth Development Strategies, Inc.

The report was completed in November 2004 and is available through Youth Development Strategies, Inc. online at:

www.ydsi.org

Youth Development Strategies, Inc.
429 Fulton Street
Philadelphia, PA 19147

Suggested Citation: Gambone, M.A., Klem, A.M., Summers, J.A., Akey, T.A., & Sipe, C.L. (2004). *Turning the Tide: The Achievements of the First Things First Education Reform in the Kansas City, Kansas Public School District.* Philadelphia: Youth Development Strategies, Inc.

About the Funders

This report is based on a project funded by the Ewing Marion Kauffman Foundation in conjunction with the Kansas City, Kansas School District and the Institute for Research and Reform in Education. The contents and views expressed herein are solely those of the authors.

This project has also been funded in part with Federal funds from the U.S. Department of Education under contract number ED-99-C0-0159. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Turning the Tide:

The Achievements of the First Things First Education Reform in the Kansas City, Kansas Public School District

Michelle Alberti Gambone, Ph.D.
Adena M. Klem, Ph.D.
Jean Ann Summers, Ph.D.
Theresa M. Akey, Ph.D.
Cynthia L. Sipe, Ph.D.

Acknowledgements

In an effort of this magnitude – seven years of planning, data collection and developing new strategies for sharing data with stakeholders in formats that are more consumable than the traditional evaluation report – there are many, many people to thank.

First and foremost we are grateful to all the staff and students in the Kansas City, Kansas school district who for five years took time out of their days to share their thoughts and opinions while enduring classroom observations, staff and student interviews, and annual survey administrations. We appreciate your honesty and hard work and admire all you have accomplished.

The Executive Committee was the initiative's leadership team and included representation from each partner organization. It was their commitment to tracking the progress of their work and their absolute commitment to accountability that made this evaluation not only feasible, but worth doing. Thank you for serving as our learning lab for understanding that more decimal places are not always better and the medium really does affect the message. Your insistence on clarity helped us develop creative forms of making complicated and important information tell a meaningful story. We (at least) will miss our two day marathon meetings.

Dr. William Moore played an integral role regardless of the hat he wore over the six years of the evaluation. His thoughtful perspective about the field of education, commitment to the project, high quality work and team building skills were critical to both getting the work done year after year and remaining more sane than not while doing so.

Special thanks to Dr. Sue C. Thompson and Dr. Loyce Caruthers, who worked with dedication to interview the faculty and students of the schools, and to help analyze factors contributing to – or impeding – reform efforts. They brought their many years of experience as educators and administrators to the analysis and were assisted by Corrin Danner Dunn and Jessica Dumas.

The quality of the data in this evaluation was dependent on the strong working partnership between the FTF research team and the KCKPS Department of Evaluation, Research, and Assessment (DERA). Dr. Dan Wright, Director of DERA, and Ms. Jill Cobb, Institutional Research Analyst, spent countless hours collecting data, preparing data sets, and cleaning data files over the last seven years. Both Dan and Jill made this research a regular part of their daily work, responding to our calls despite the fact that they knew it meant adding work to their already full plates and striving to improve the quality of data collected throughout the district.

Several other members of the research team made irreplaceable contributions to this work.

- Donna Sulak has earned a thousand thanks for her work with the KCKPS datasets over the course of the evaluation. After six years of carefully cleaning, documenting, and analyzing KCK data, she has officially earned the title of "the most careful data analyst ever."
- Drs. Nona Tollefson and Steve Lee at the University of Kansas were instrumental in developing the classroom observation strategy.
- Dr. Phyllis Clay trained and oversaw the classroom observers and managed the FTF survey administration. Our response rates would have been much lower without her!

We'd like to express our gratitude to colleagues who wrangled their schedules and commitments in order to provide guidance and advice on the analytic design of this complex evaluation. Dr. Mark Berends (Vanderbilt University), Dr. Ann Blanc (Blancroft Research International), Dr. David Francis (University of Houston), and Dr. Richard Serpe (California State University, San Marcos) reviewed our analytic methods and provided invaluable recommendations for how to maintain the highest levels of scientific rigor while maximizing the utility of the results to the field.

Dr. Berends, Dr. Blanc, Stacey Daraio (Community Network for Youth Development), and Dr. Adam Gamoran (University of Wisconsin, Madison) dedicated hours to reviewing drafts of the report and provided essential feedback on key issues in the fields of education, youth development and methodological rigor.

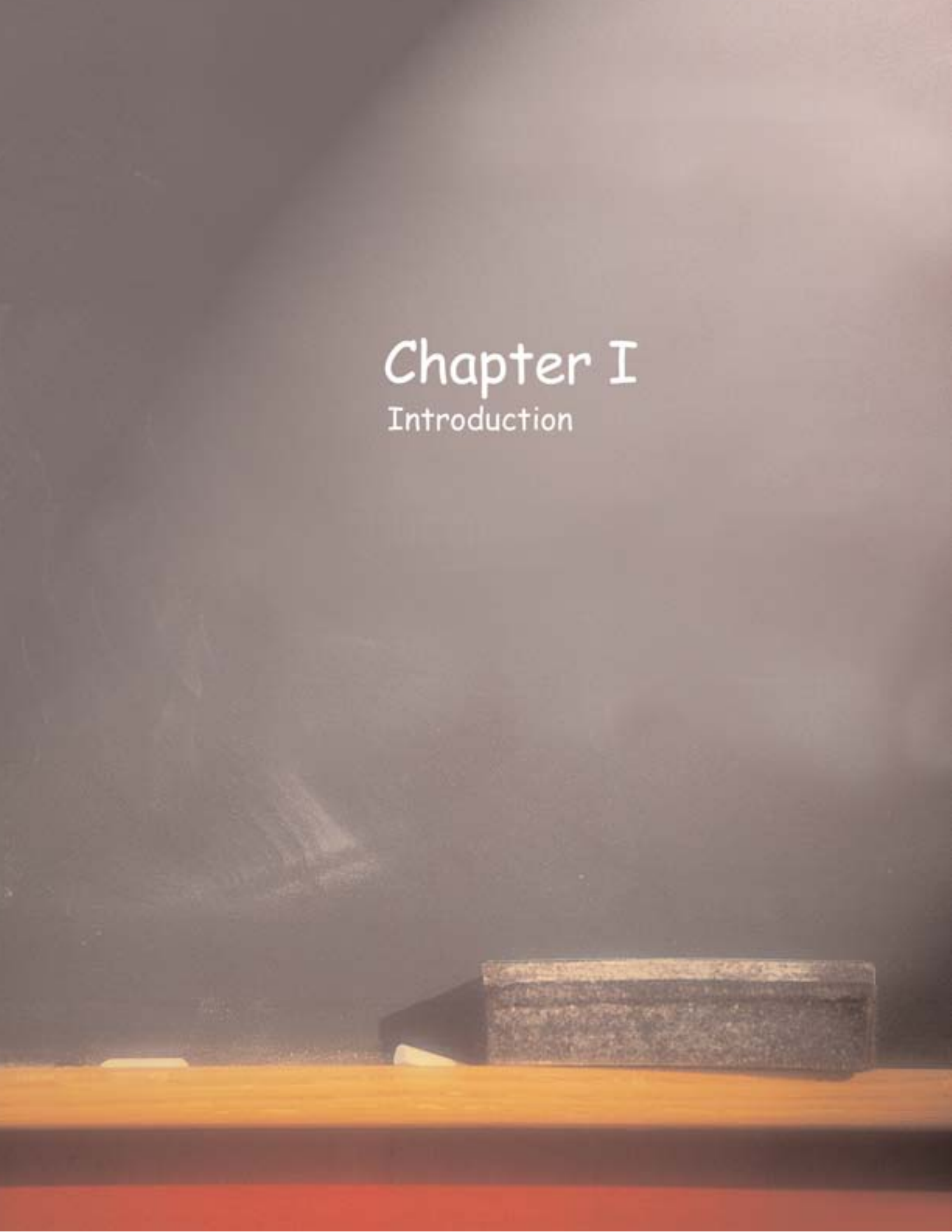
Special thanks to Dr. Blanc who not only provided expertise in log-odds models and survival analysis, but who also persevered with us in our continuing pains to present results in a format meaningful to diverse audiences.

Drs. Howard Bloom, Fred Doolittle and Janet Quint from MDRC also provided invaluable insight as we identified the best strategies for analyzing, presenting, and discussing the results of this evaluation.

We are indebted to the editorial team - Natalie Jaffe, Maxine Sherman and Lesley Dahlkemper for their outstanding skills in pushing us to do the best we could to boil seven years of intensive implementation and masses of data into an intelligible story. Our thanks also to Janice McLemore Graphic Design.

Finally, a very special thanks to Jan Nevin who invested hours of effort and creative energy into helping us develop strategies for presenting results to stakeholders in the district while the reform was underway. Jan excels at finding new ways to present information to people in meaningful ways – allowing school and central office staff, researchers and funders to see a clear picture of what the data were showing as the initiative unfolded.

TABLE OF CONTENTS



Chapter I

Introduction

INTRODUCTION

In 1996, the Kansas City, Kansas, Public School (KCKPS) District decided to implement a comprehensive school reform model – First Things First – district-wide. Unlike most districts across the country, individual schools were not given the choice of selecting their own reform model; instead, KCKPS leaders were determined to devote all of their resources to making a single vision a reality. They believed that if each school worked as an individual unit with its own set of requirements, the central office's approach to improvement would be scattered and, thus, ineffective.

Kansas City had learned from the experiences of other districts that pursued reform models during the 1990s with support from the federal government and private foundations. Individual schools were allowed to select from a range of models (as many as five or more in each district at any one time), with unsatisfactory results that caused a shift away from school-level reform and toward district-level reform (Berends, Chun, Shuyler, Stockley & Briggs, 2002; Slavin, 2003, 2004). In the wake of this change, the field of education has begun exploring the linkages between systemic reform and student performance.

KCKPS is a medium-sized urban school district of approximately 21,000 students, of whom 79 percent are minority and 74 percent are eligible for free and reduced-price lunch. Students experience most of the problems that plague urban schools, from coming to school hungry to exposure to drugs and violence.

This report charts First Things First's policy context, history, implementation success, intermediate outcomes, student achievement outcomes and critical issues. It examines whether it is possible to implement an externally developed comprehensive reform model in an entire school district; and if so, whether and how much student outcomes improve. We found that implementation of such a model is indeed possible and that student outcomes did improve significantly.

WHAT DID THE REFORM PROCESS LOOK LIKE IN KANSAS CITY, KANSAS?

First Things First (FTF) is a K-12 education reform initiative designed to raise the academic performance of all students to levels required for post-secondary education and high-quality employment. The reform began in Kansas City, Kansas, with the creation of a partnership between the school district, the reform model's developer and the Ewing Marion Kauffman Foundation. After a year of planning for change at the district level from May 1996 to May 1997, the reform started in Kansas City with a reshaping of the responsibilities of central office staff, a reallocation of district resources, and a phasing in of a planning and implementation process in the school buildings.

FTF is premised upon the need for strong, supportive relationships between teachers and students, and among staff at all levels; and effective instructional practices that engage students in rigorous and meaningful academic content. These conditions are the mechanism for achieving better student outcomes. The model identifies seven critical features of school reform (detailed below) that will allow schools to put these conditions into place. In August 1997, one-quarter of the district's schools began creating plans to put these critical features in place for the following school year.

DID OUTCOMES IMPROVE IN KANSAS CITY, KANSAS?

The question of primary importance is whether anything changed in the KCKPS District since 1996 and the implementation of the FTF reform? Simply put, is it worthwhile to follow the path of reform in this district – first and foremost, because the raw data show outcomes for students significantly improved over this period. This section shows the raw trends in selected areas over the course of the reform

effort – with no statistical controls or complicated models. The remainder of the report then turns to examining how these improvements were achieved and the extent to which these improvements can be attributed to the strategies associated with the FTF reform.

Key outcomes were tracked over the course of the initiative to guide the work of the FTF partnership. Each year, they examined data to see whether:

- students were coming to school more regularly;
- students' relationships with teachers were better;
- students were more engaged in school;
- student achievement in reading and math was improving; and
- high school students were less likely to drop out and more likely to graduate.

Are more students coming to school? Table I-1 shows the change in the building attendance rates for elementary, middle and high school students in KCKPS between the 1997-1998 school year and the 2002-2003 school year. Elementary grade rates stayed consistently high (about 95%), as occurs in most schools serving lower grade levels. Within the secondary schools, the number of students attending school steadily increased, with the exception of the 2002-2003 school year, when a slight dip in attendance rates occurred in high schools.

Are more students experiencing good relationships with teachers? Supportive relationships between teachers and students are essential ingredients for improving performance. As discussed in greater detail below, surveys were administered to students in grades 3 through 12, measuring students' experience of support from their teachers. The six-year trend in Table I-2 shows marked improvements across all grade levels in the percentage of students experiencing high-quality relationships with their teachers, particularly at the middle and high school levels where about double the number of students had good relationships relative to pre-FTF implementation.

Table I-1: Average Building Attendance Rates

	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
Elementary School	95%	94%	95%	94%	97%	95%
Middle School	88%	90%	89%	94%	95%	94%
High School	77%	82%	84%	86%	90%	87%

Table I-2: Percentage of Students Experiencing High Levels of Teacher Support

	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
Elementary School	20%	21%	22%	32%	33%	38%
Middle School	22%	27%	31%	35%	38%	42%
High School	17%	20%	20%	27%	27%	35%

Table I-3: Percentage of Students Experiencing Low Levels of Engagement

	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
Elementary School	33%	32%	32%	28%	26%	31%
Middle School	45%	39%	38%	28%	26%	16%
High School	47%	47%	44%	33%	32%	24%

Are fewer students disengaged from school? Research has shown a strong link between the level of supportive relationships students experience and their engagement in the classroom (Connell and Wellborn, 1991; Marks, 2000; Ryan & Patrick, 2001; Skinner and Belmont, 1993; Solomon, Battistich, Watson, Schaps & Lewis, 2000; Voelkl, 1995). In turn, engagement has been shown to predict improvements in student performance (Lee & Smith, 1993, 1995; Roderick & Engel, 2001; Willingham, Pollack, & Lewis, 2002). The student surveys administered in the district measured levels of student engagement. Table I-3 shows a substantial decrease in the percentage of students who had very low levels of engagement – that is, who were disengaged from school. Secondary students, particularly high school students, who were the most likely to be disengaged from school, showed the greatest improvements during the study period.

Are high school students less likely to drop out and more likely to graduate? Across the four comprehensive high schools in the district, the average building dropout rate steadily declined since the implementation of FTF (see Table I-4).

During the same time period, the district-wide graduation rate declined from 57 percent to 48 percent until school year 2000-2001, when the district-wide graduation rate showed a substantial increase from 48 percent to 69 percent.

Has student achievement in reading and math improved? On the Kansas State Reading and Math tests, students score in one of five proficiency levels: unsatisfactory, basic, proficient, advanced and exemplary.[1] The district uses two standards to assess whether student performance is improving on each of these tests. One tracks whether the number of students performing at, or above, the satisfactory level (satisfactory, proficient or advanced) is increasing. The other assesses whether the number of students performing at the lowest level (unsatisfactory) is decreasing. The reading portion of the test is administered in grades 5, 8 and 11; the math portion in grades 4, 7 and 10. The test is administered to all

Table I-4: Building Dropout and Graduation Rates

	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
Dropout Rate	11%	13%	8%	8%	5%	5%
Graduation Rate	57%	54%	48%	61%	61%	69%

Table I-5: State Reading Test Scores

	2000-01	2001-02	2002-03
Elementary % Proficient or above	29%	32%	47%
Elementary % Unsatisfactory	41%	37%	21%
Middle School % Proficient or above	36%	37%	56%
Middle School % Unsatisfactory	33%	29%	14%
High School % Proficient or above	25%	29%	34%
High School % Unsatisfactory	45%	44%	36%

[1] See the Kansas State Department of Education website for additional information on Kansas Assessments: www.ksbe.state.ks.us.

students in these grades across the state, including special education students and English Language Learner (ELL) students.[2] Tables I-5 and I-6 show the change in the percentage of students scoring at or above proficiency and those scoring at unsatisfactory levels on the Kansas State Reading and Math Tests. District students at all three levels illustrate movement out of the unsatisfactory level and into the proficient or above categories on the reading test.

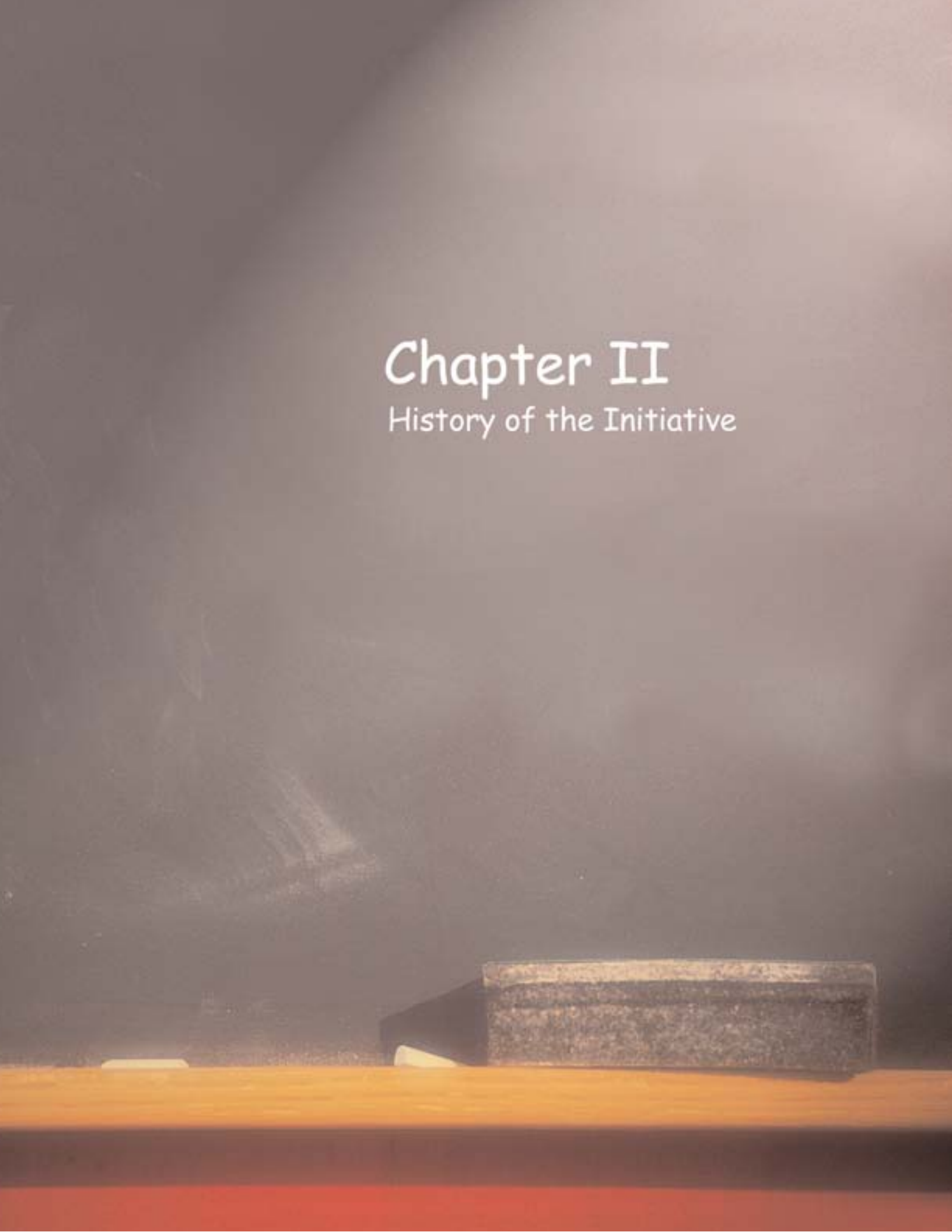
Though high school students showed little movement on the math portion of the state test, elementary and middle school students demonstrated considerable improvement between 2001 and 2003.

The raw data show that outcomes have improved in the district. More students are attending school, have better relationships with their teachers, are performing better on the state achievement test and are graduating from high school. At the same time, fewer students are disengaged or dropping out of school. The question at hand is whether these changes are meaningful and, if so, whether they can be attributed to implementation of First Things First. The remainder of this report addresses these questions.

Table I-6: State Math Test Scores

	2000-01	2001-02	2002-03
Elementary % Proficient or above	31%	33%	43%
Elementary % Unsatisfactory	40%	34%	28%
Middle School % Proficient or above	16%	19%	24%
Middle School % Unsatisfactory	61%	56%	50%
High School % Proficient or above	14%	15%	16%
High School % Unsatisfactory	63%	57%	62%

[2] Unlike the other measures reported here, the state test scores can only be reported for years 2000-2001 through 2002-2003 because the state revised the test in 2000-2001 and the revised test is not comparable with the previous version.



Chapter II

History of the Initiative

POLICY CONTEXT FOR COMPREHENSIVE SCHOOL REFORM[3]

From the mid-1960s through the early 1990s, most of the programs that the federal Title I program for disadvantaged students provided to schools were piecemeal and uncoordinated. Disappointed by the results, policymakers opted for "comprehensive school reforms" that encouraged school-wide improvement. The 1994 reauthorization of the federal Elementary and Secondary Education Act encouraged schools where at least half of the students were disadvantaged to implement these school-wide reforms. Unlike past initiatives, comprehensive school reform models ask schools to overhaul themselves from top to bottom by using a single school-wide vision as a focus for redesigning curriculum, student assessment, professional development, governance, management and other key functions (American Institutes for Research, 1999; Berends, Kirby, Naftel & McKelvey, 2001). These models were designed to help individual schools improve their student outcomes and, while they were based on existing research about best practices, few of the models had a strong research base about the effects on student achievement (American Institutes for Research, 1999; Borman, Hewes, Overman & Brown, 2003).

Between the Title I program's continuing focus on school-wide change and the efforts of independent comprehensive school reform model developers, states and districts began to encourage – or require – their schools to implement school-wide models. For instance, Memphis, Tennessee, required all 164 of its schools to adopt an improvement program of their own choice in 1995. Each school adopted a school design – many from the New American Schools – and each school received some of its support services from the model providers. Other districts attempted similar initiatives to adopt well-known, external reform models on a large scale, including San Antonio and Miami-Dade County.

However, by the late 1990s, districts that enthusiastically embraced comprehensive school reform found it difficult to manage large numbers of very different reforms chosen by different schools. Disillusioned, many of these districts abandoned their efforts or scaled them back (Berends, Chun, Shuyler, Stockley & Briggs, 2002; Slavin, 2003; Slavin & Madden, 2004).

About the same time that districts were beginning to scale back their implementation of comprehensive school reform models, the education policy environment changed with the passing of the 2001 No Child Left Behind Act (NCLB). NCLB contains the stringent accountability requirement that every student in the U.S. – including minorities, low socio-economic status (SES) students, special education students, and English Language Learners – is expected to demonstrate proficiency on state reading and math tests by 2014. The Act requires states to annually test their students and have a certain percentage of students achieve proficiency each year. Schools and districts that do not meet these targets for consecutive years face severe consequences.

As one of the strategies for helping states meet these accountability targets, NCLB incorporated comprehensive school reform components directly into the Title I program. Under NCLB, schools identified as needing improvement must pursue strategies designed to improve achievement, including comprehensive school reform.

[3] The contents of this chapter and the next are adapted from Gambone, Klem, Moore & Summers, 2002

THE SHIFT TOWARD DISTRICT-WIDE REFORM

The disillusionment experienced by districts implementing comprehensive school reform models on a school-by-school basis, combined with the rigorous testing demands of NCLB, caused many districts to turn toward district-wide strategies (Viadero, 2004). At the same time, the education field began to explore the outcomes associated with district- and even state-level reform in greater detail.

For instance, a national task force convened by the Annenberg Institute for School Reform in 2000 began tackling the question of how to redesign districts so that large numbers of high-performing schools could flourish (Olson, 2000). The taskforce focused on the key supports and services districts provide to schools and how those supports need to be rethought in light of standards-based reform and NCLB (Kronley & Handley, 2003).

Research on the role of the district in educational change was also supported by both public and private funders. This research identified several key elements that help high-poverty districts improve their students' performance (Togneri & Anderson, 2003; McLaughlin & Talbert, 2003; Massell and Goertz, 2002; Cawelti & Protheroe, 2001; Snipes, Doolittle & Herlihy, 2002; Skrla, Scheurich, & Johnson, 2000). First and foremost, successful districts have leaders who publicly acknowledge poor student performance, and who demonstrate commitment and support to helping schools ensure that all students are able to meet state standards. These leaders make it clear that they are in the improvement process for the long haul – the reform will not simply go away, so there is no point in waiting for it to do so.

Other common elements of successful high-poverty districts include a system-wide vision; more time built into school schedules for staff to work together; data-driven decision-making; a shift in the role played by central office staff, from control and supervision to support for instructional improvement; and targeted professional development around those instructional improvement efforts.

The general conclusion reached by this research is that "if low-performing schools are to be transformed into high-performing learning communities, the total system must be considered as changes are made. Failure to make systemic changes has been a major flaw in prior attempts to improve student achievement, which focused on individual school improvement" (Cawelti, 2003, p.1). It appears that when a reform effort is systemic (with the whole district supporting a coherent reform vision), the district is far more likely to be successful in improving student achievement (Snipes, Doolittle & Herlihy, 2002; Cawelti & Protheroe, 2001).

MAKING SYSTEMIC REFORM A REALITY

Research increasingly shows that systemic reform can be implemented and improve student achievement. Yet, not very many high-poverty districts are engaging in this type of system-wide change.

Part of the challenge is that districts struggle with choosing among the plethora of reform strategies available to help improve student performance. There are 30 comprehensive reform models and 10 additional reading/language arts programs that have been accepted by the Catalog of Comprehensive School Reform as demonstrating evidence of effectiveness in improving student academic achievement, extent of replication, implementation assistance provided to schools and comprehensiveness. Hundreds of other national and local reform initiatives are available to districts that were not included in the catalog.

And while many comprehensive school reform models have evidence supporting the link between implementation of the model and improved student outcomes, few have conducted research for all of the schools in a district. Of the sparse number of research studies exploring district-wide reforms, none can demonstrate an empirical link from district policies and practices to changes in teaching and learning practices to outcomes at the classroom or building level (Anderson, 2003). Thus, districts often take a leap into the unknown when selecting a district-wide reform initiative.

THE KANSAS CITY, KANSAS SCHOOL DISTRICT

At the onset of the reform in 1996, the Kansas City, Kansas, School District was a largely urban district serving approximately 21,000 students in 47 buildings. The geography of the district places its eastern boundary directly across the Missouri River from Kansas City, Missouri. The District has a marked economic boundary as well: households in the

western reaches of the District are more affluent than those in the eastern and northern segments of the District. When the District adopted FTF as the key reform strategy, 68 percent of its students received federally subsidized lunches and about 70 percent of the students were of non-white ethnic groups. Average daily attendance was roughly 90 percent. Performance on standardized achievement tests was substantially below the national norm and markedly substandard on state standards-based assessments in all four core curricular areas – reading, math, social studies and science. (See Table II-1 for a summary of the District's characteristics between 1996 and 2003.)

During the first three years of the initiative, the size and demography of the student population changed most notably in the areas of ethnicity, test scores and enrollment. By 1999, the percentage of minority students increased to 74 percent, principally due to increasing Hispanic enrollment and declining white student enrollment. There was also a steady increase in the percentage of students scoring in the bottom quartile of the standardized reading achievement test administered in the district. Finally, the 1999-2000 academic year showed District enrollment stabilizing for the first time in three decades.

Table II-1: District Demographics

District Characteristics	1996	1997	1998	1999	2000	2001	2002	2003
Number of Schools (High/Middle/Elementary)	6/8/33	5/8/31	5/8/30	5/8/30	5/8/28	5/8/28	5/8/28	5/8/28
Total Enrollment	21670	21233	21249	20926	20726	21173	21215	20775
% Subsidized Lunch	64%	65%	66%	66%	68%	71%	72%	74%
% Minority	68%	69%	70%	72%	74%	76%	77%	78%
Average Daily Attendance	91%	91%	88%	90%	91%	93%	94%	93%

** Data from Kansas State website.*

INTRODUCING FIRST THINGS FIRST TO KANSAS CITY, KANSAS

The partnership to undertake the FTF initiative in Kansas City, Kansas (KCK) was formed in early 1996. The Ewing Marion Kauffman Foundation (Kauffman) – a Kansas City, Missouri-based, national foundation supporting youth initiatives – became familiar with the Institute for Research and Reform in Education's (IRRE) education reform framework and introduced its plan for a district-wide effort to the KCK school district leadership. In the course of its work, IRRE – a Philadelphia-based, non-profit intermediary working on youth development efforts – developed a process for introducing and explaining the FTF theory of change to various education stakeholders. This familiarization process, called "Roundtables," was designed to present the FTF theory of change by mapping the pathway to long-term outcomes; focusing on the critical features of necessary school site reform; and exposing participants to the realities of putting the critical features in place through presentations from school administrators, teachers and students whose schools had implemented these practices.

The KCK district leadership was invited by Kauffman in May 1996 to attend an FTF Roundtable. In an attempt to reverse the negative trend of students achieving well below the national and state averages for several years, the KCK superintendent had recently completed a series of efforts to create a systematic, data-driven planning and evaluation process. Additional efforts were being developed to enhance curriculum and instruction. After attending the Roundtable, the KCK district leadership believed that FTF could serve as an effective vehicle to synthesize their efforts toward improvement. Kauffman agreed to entertain a joint proposal from the district and IRRE to implement FTF in Kansas City, Kansas.

During the following summer, IRRE worked with the KCK district leadership to draft an Accountability Plan for implementing FTF in KCK, which laid out in detail the actions to be taken and the responsibilities

of each partner. Meetings were held between Kauffman, the district, IRRE and School Board staff to review and discuss the plan. The product of these meetings was an agreement by the three partners to launch FTF in Fall 1996.

PREPARING FOR CHANGE AND RESTRUCTURING CENTRAL OFFICE RESOURCES

Between May 1996 and May 1997, District leadership participated in restructuring and planning activities to prepare for implementation of the FTF initiative. In October 1996, a second Roundtable was held to present FTF to the KCK Board of Education. Subsequent conversations led to Board of Education approval and the submission of a joint proposal from the District and IRRE to the Kauffman Board in November 1996. This proposal included two key components of the FTF initiative in KCK.

Phase-In Plan

The district proposed to gradually phase schools into the initiative in clusters – a high school and its feeder elementary and middle schools – to ensure that the necessary resources would be available for schools as the central office began reallocating its personnel and funding to support the planned changes. The first and second clusters of schools chosen for FTF implementation were selected because of their differences on key indicators, such as drop-out rates, graduation rates, daily attendance and student demographics. Wyandotte was identified as the first cluster to begin planning for FTF implementation because the high school had the poorest performance profile of the high schools in the district. Compared with the other four high schools, Wyandotte had a lower graduation rate (53%) and average daily attendance than the others (< 75%). It also had a predominately minority student population (82%) with almost 75 percent of its students receiving subsidized lunches.

The Washington cluster was identified as the second group of schools to be phased into FTF because its population was the most dissimilar from the Wyandotte cluster. Washington High had a 76 percent graduation rate, an average daily attendance rate of 90 percent and a 58 percent minority student population, with 41 percent of students receiving subsidized lunches. The socioeconomic difference between the Wyandotte and Washington clusters – one urban core with high poverty and the other more suburban-like with relatively low poverty rates – was seen as an opportunity to learn about the process of systemic change in two areas of the community with very different histories, needs and challenges. District leaders believed that if FTF could be successfully implemented in these two very different clusters it would demonstrate that it could also be implemented in the remaining schools. The final two clusters – Harmon and Schlagle – were to be phased in one at a time in subsequent years.

In fact, the pace of implementation was accelerated so that the Harmon and Schlagle clusters both started their reform work in the 1999-2000 school year. This change was made in response to events in the clusters. In anticipation of implementation, staff in some buildings had begun to implement what they believed the reform strategies were before actually participating in any of the initiative's educational and planning sessions. In order to forestall the need to have schools "undo" what they believed were reforms that met the District's goals, the decision was made to bring them into implementation together in 1999.

Schools in each cluster spent one year creating plans to implement the FTF critical features and began implementation the following year. The Wyandotte cluster began its planning year in the 1997-1998 school year; Washington entered its planning year in 1998-1999. The final two clusters began planning in the 1999-2000 school year.

School Improvement Facilitators (SIFs)

The administration reassigned district-level curriculum specialists to the position of SIFs – leaders of the change process in schools. This reallocation of positions created the necessary building-level support to do site planning and begin implementation. SIFs required training in system- and building-wide change, facilitation skills, team building, and effective communication strategies. Concurrent with this reallocation came a streamlined curriculum department, although most SIFs maintained duties as both curriculum specialists and SIFs.

With approval of the linked proposals from the Kauffman Board, the partners – the KCK District, IRRE and Kauffman – developed a three-way accountability plan that included the creation of two managing bodies: the Executive Committee (EC) and Research Management Team (RMT) to oversee implementation and evaluation, respectively (see

Text Box A for a description of these bodies). It also delineated the activities each partner agreed to accomplish over the time period covered in the plan, and spelled out the consequences associated with not completing tasks (e.g., suspension of funding). The plan outlined in great detail each action step to be taken and identified the partner responsible for ensuring the completion of each step. These steps were arrived at through a joint decision-making process among the three partners. These plans were updated annually with new tasks and responsibilities.

This clarification of roles and responsibilities in the accountability plan helped maintain external pressure on all three partners – pressure that became important for the continuation of FTF in KCK. The plans helped the District develop a sense of accountability for building capacity for change, as well as for improving test scores. The funder not only provided money but also participated in major events, provided in-house support from foundation departments (e.g., research and communications), and responded to the ongoing needs of the district. Finally, IRRE became much more than the "outside expert" serving as a technical assistance provider. It worked closely with all levels of the district from advising the superintendent to helping school staff plan and implement the critical features. As designer of FTF, IRRE played the key role of monitoring the fidelity of local planning and implementation, and became a sounding board for local decisions that were possible distractions from the main work of the reform. For example, when opportunities to apply for grants and other sources of funding arose, IRRE staff urged the District to ask, "Will this help move FTF forward or will it create additional work that doesn't fit with our vision of reform?" Over time, district leaders began naturally to ask that question of themselves and their partners.

As the initiative unfolded, the leadership continued to develop strategies to support the buildings in their efforts. A timeline of the main activities at the district level and brief descriptions of the implementation strategies used by the schools are included in Chapter V. The next chapter outlines the reform model and is followed by a description of the research design.

TEXTBOX A

First Things First Managing Bodies in Kansas City, Kansas

Executive Committee

Roles and Responsibilities

The purpose of the FTF Executive Committee is to make strategic decisions about FTF implementation, monitor progress of the initiative and create operational plans to address barriers to successful implementation across all district schools. In addition, the Executive Committee meets regularly with researchers to examine trends in implementation and to request information to support their own decision-making.

Members of the Executive Committee

Fall 1996: In conjunction with the early informal collaboration of the partnership (the District, IRRE and Kauffman), the first iteration of the FTF Executive Committee emerged, consisting of:

- District Associate Superintendent;
- President of IRRE; and
- Senior Program Officer at Kauffman

Over time, the core leadership group expanded to include the Superintendent, FTF Director of School Improvement, two Executive Directors of School Operation, and Superintendent's Management Team, representing Professional Development, Research and Assessment, Special Education, Instructional Development, and Curriculum and Standards.

Research Management Team

Roles and Responsibilities

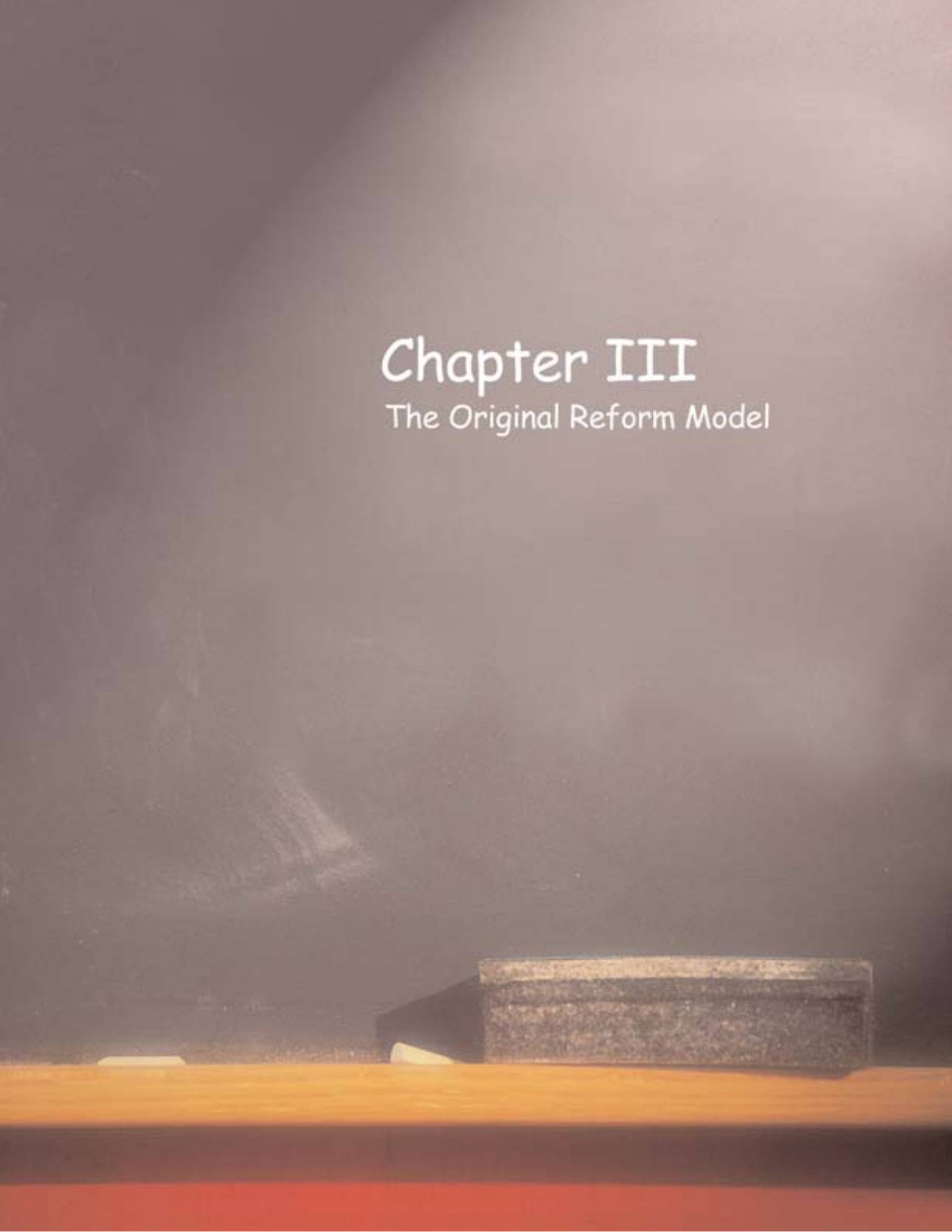
The purpose of the Research Management Team is to design, manage and disseminate findings regarding the implementation and effects of FTF in the Kansas City, Kansas Public Schools.

Responsibilities of the Research Management Team include:

- Contract with independent researchers to conduct the studies necessary to document implementation and effects of FTF;
- Advise the FTF Executive Committee;
- Provide additional support for data collection and analysis conducted by the District's Research and Assessment Department (with the help of their consultants); and
- Prepare all official reports documenting the course of FTF in the district.

Members of the Research Management Team

Spring 1997: Shortly after the establishment of the FTF Executive Committee and upon their direction, an independent research consultant was hired to coordinate the research and evaluation activities associated with FTF. The president of Youth Development Strategies, Inc. (YDSI), a Philadelphia-based organization, was selected. One senior research associate from Kauffman and one from the district served with the president of YDSI as the Research Management Team.



Chapter III

The Original Reform Model

It was within this context that the FTF initiative was undertaken in Kansas City, Kansas (KCK) in 1996. The Institute for Research and Reform in Education (IRRE) combined into a single, comprehensive model its own and others' research on the essential features of whole-school reform, research on organizational change, research on youth development, and a plan for initiating and supporting change through district-wide restructuring and realignment of resources. This model – or "theory of change" – was used from the outset of the initiative to garner support from, and guide the activities of, all stakeholders, including the funders, district leadership, school board, school-building administrators, school-building staff and the community.

Because of the centrality of this theory of change to both the initiative and the research design used to evaluate it, the next section briefly lays out the logic of the framework and identifies the elements that are evaluated in this report.

THE FIRST THINGS FIRST FRAMEWORK[5]

The FTF theory of change (shown in Figure III-1) proposes a set of early, intermediate and long-term changes needed to produce significant system-wide improvement in student outcomes. Starting with the longer-term outcomes desired for youth, the model works backward to the developmental milestones in education needed to achieve these outcomes, outlines the supports and experiences required to achieve these milestones, describes the school-building restructuring necessary to ensure these supports are in place for both students and adults, and finally outlines the district-level activities required to create the conditions and capacity for system-wide change. Figure III-1 illustrates the key elements and outcomes associated with each of these steps, which are briefly described below.

Box A. What are the long-term goals for youth?

The long-term outcomes – decent jobs, good relationships and the ability to contribute to the community in positive ways (Box A, Figure III-1) – are policy goals we care most about. These are also the outcomes that require earlier accomplishments in the school-aged years (Brown & Emig, 1999; Halperin, 1998; Hauser & Sweeney, 1997; Maggs, Frome, Eccles & Barber, 1997; Plank & Jordan, 1997).

Box B. What educational outcomes lead to these long-term goals?

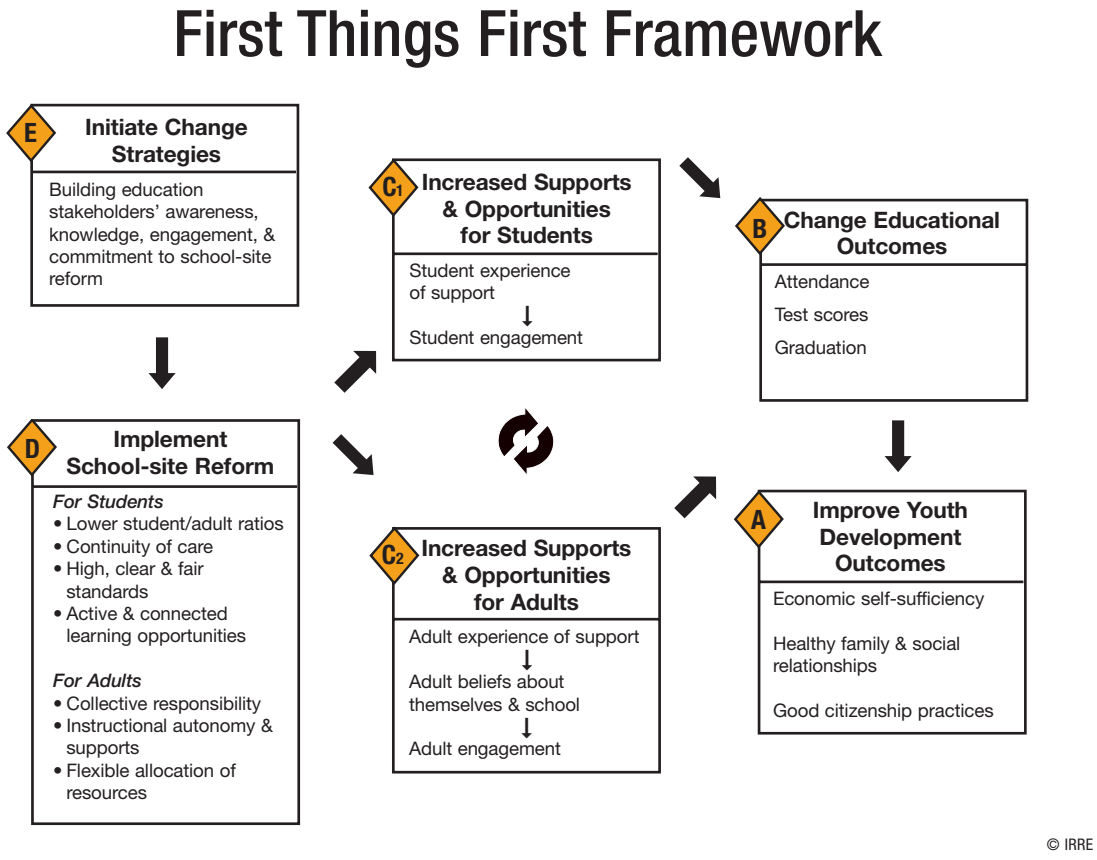
Longitudinal research has shown that children must master the ability to be productive in order to achieve later success in life (Brown & Emig, 1999; Hauser & Sweeney, 1997; Maggs, Frome, Eccles & Barber, 1997; Padilla, 1997; Plank & Jordan, 1997). During the school years, the two markers of this developmental milestone that most strongly predict the adult outcomes in the framework are (1) how well students do in school academically (e.g., performance on standardized tests, grades), and (2) how committed they are to their education (e.g., attendance, suspensions, expulsions) (Box B, Figure III-1). These outcomes are included as the focus of educational systems undergoing change because of their association with later success.

Box C. What supports and opportunities improve educational outcomes?

According to the framework, significant changes in the education environment are required in order to

[5] The version of the FTF framework presented here is the one used for the first three years of work in KCK. As the initiative unfolded, the framework has been revised to reflect feedback from the progress of the effort. This type of modification is a hallmark of using a "theory of change" process to guide an initiative. For a description of the revised version of the framework see *First Things First, A Framework for Successful School-Site Reform*, 2003. For a description of the theory of change process see *You Can Get There From Here: Using A Theory of Change Approach to Plan Urban Education Reform*, 2000.

FIGURE III-1



improve student performance and commitment. Specifically, changes must occur in the everyday lives of students in their classrooms and schools. Students should experience better interpersonal and instructional supports, which in turn lead to more positive beliefs and greater engagement in school (Box C_1, Figure III-1). For these changes to occur for students, schools and districts also need to simultaneously increase the supports and opportunities for the adults in schools, which lead to more positive beliefs and greater engagement on the part of these adults (Box C_2, Figure III-1). The theory holds that as the experience of these supports and opportunities are strengthened, educational outcomes can be expected to improve.

Box D. How should schools change in order to increase supports and opportunities?

In order to improve the teaching and learning environment of all classrooms, the framework calls for three types of changes in schools. These changes are referred to as First Things First's "seven critical features" of school-site reform (Box D, Figure III-1). These critical features – four for students and three for adults – provide the parameters for change activities that are called for in the framework to better support youth and adults. They represent the intermediate outcomes that, if achieved, signal progress toward improving educational supports and opportunities and, ultimately, toward achieving the desired educational and long-term outcomes for youth. They are listed here because of their

centrality to all of the consensus building, capacity building and planning activities of the initiative that are the focus of this report.

SEVEN CRITICAL FEATURES OF SCHOOL-SITE REFORM[6]

FOR STUDENTS:

Structural Critical Features

1. Lower student/adult ratios to no more than 15 to 1 during instruction in core academic subjects (reading and math) through redistribution of professional staff.

Implementing this critical feature will require schools to consider how to reorganize schedules and staffing. For instance, a school might pull out rotating groups of students to attend elective courses while the remaining students participate in reading or math instruction; and/or schools might train special subject staff (art, music, physical education), paraprofessionals, and aides to teach reading and/or math.

2. Provide continuity of care by having the same group of 8 to 10 professional adults within each school level stay with the same group of no more than 120 students for extended periods of time during the school day, for at least three years in elementary school, all three years of middle school and at least two years in high school.

Implementing this critical feature requires decisions about how to restructure the school. Options for continuity of care across school years include establishing Small Learning Communities (SLCs),[7] multiage groupings,[8] and/or looping.[9] Options for continuity of care across the school day typically involve some form of block scheduling, in which classes last for longer periods of time than the traditional 48-minute class.

Instructional Critical Features

3. Set high, clear, and fair academic and conduct standards. Academic standards define what all students will know and be able to do within and across key content areas by the time they leave high school and at points along the way in their school career. Conduct standards define how adults and students should behave and; are agreed upon by adults and students; are reinforced by adults modeling positive social behaviors and attitudes; are sustained by clear benefits and consequences.

Implementing academic standards includes making decisions about how to align district, state and national standards so that students are successful on all three assessments; and integrate performance standards into everyday instruction. Implementing conduct standards includes developing a protocol for identifying student and staff agreed-upon standards for behavior for all people in the building; and developing a system for identifying rewards and consequences.

4. Provide enriched and diverse opportunities:

- To learn, by making learning more authentic (active, cooperative, integrated and real-world based);
- To perform, by utilizing assessment strategies linked directly to standards that use multiple modes of learning and performance; and
- To be recognized, by creating individual and collective incentives for student achievement, as well as leadership opportunities in academic and non-academic areas.

FOR ADULTS:

5. Assure collective responsibility, by providing collective incentives and consequences for teaching teams and schools based on improvement in student performance.

---

[6] The following list of critical features reflects the original version used in KCK. The critical features have since been revised to reflect lessons learned in KCK.

[7] Small Learning Communities (SLCs) are also known as "schools-within-a school," "houses" or "families." Each SLC has its own group of teachers and students; and sometimes its own physical space within the school, governance system and budget.

[8] Multi-age classrooms are created by combining students from different grade levels in one class regardless of age.

[9] Looping requires a teacher or team of teachers to teach the same group of students for multiple grade levels (e.g., sixth, seventh and eighth) and multiple years (between two and four).

Implementing this critical feature requires staff to decide on annual targets for student performance; to establish procedures for deciding how to establish those targets; and to decide on the incentives and consequences associated with meeting or not meeting those targets. To share responsibility for meeting targets, staff need to have scheduled time together to reflect on student performance and work on improving instruction. Establishing a common daily planning time and common professional development activities during and after school hours are two techniques for doing so.

6. Provide instructional autonomy and supports[10] to these teams of teachers so that they can develop instructional strategies that will best meet the individual and collective needs of their students.

Implementing this critical feature involves deciding the level at which decisions about instructional practice and professional development should be made – within Small Learning Communities (SLCs), grade-level committees, etc. Some issues to consider include: deciding what instructional strategies to use to support students' learning; how to obtain ongoing data on students; looking at student performance to study the effects of teaching; and how to sustain this repertoire of instructional strategies. Strategies for improving instruction include using common planning time and professional development activities.

7. Allow for flexible allocation of available resources by teams and schools, based on instructional and interpersonal needs of students. Resources include **people** (students and staff); instructional **facilities** (on and off campus); instructional, planning and professional development **time**; and discretionary **funds**.

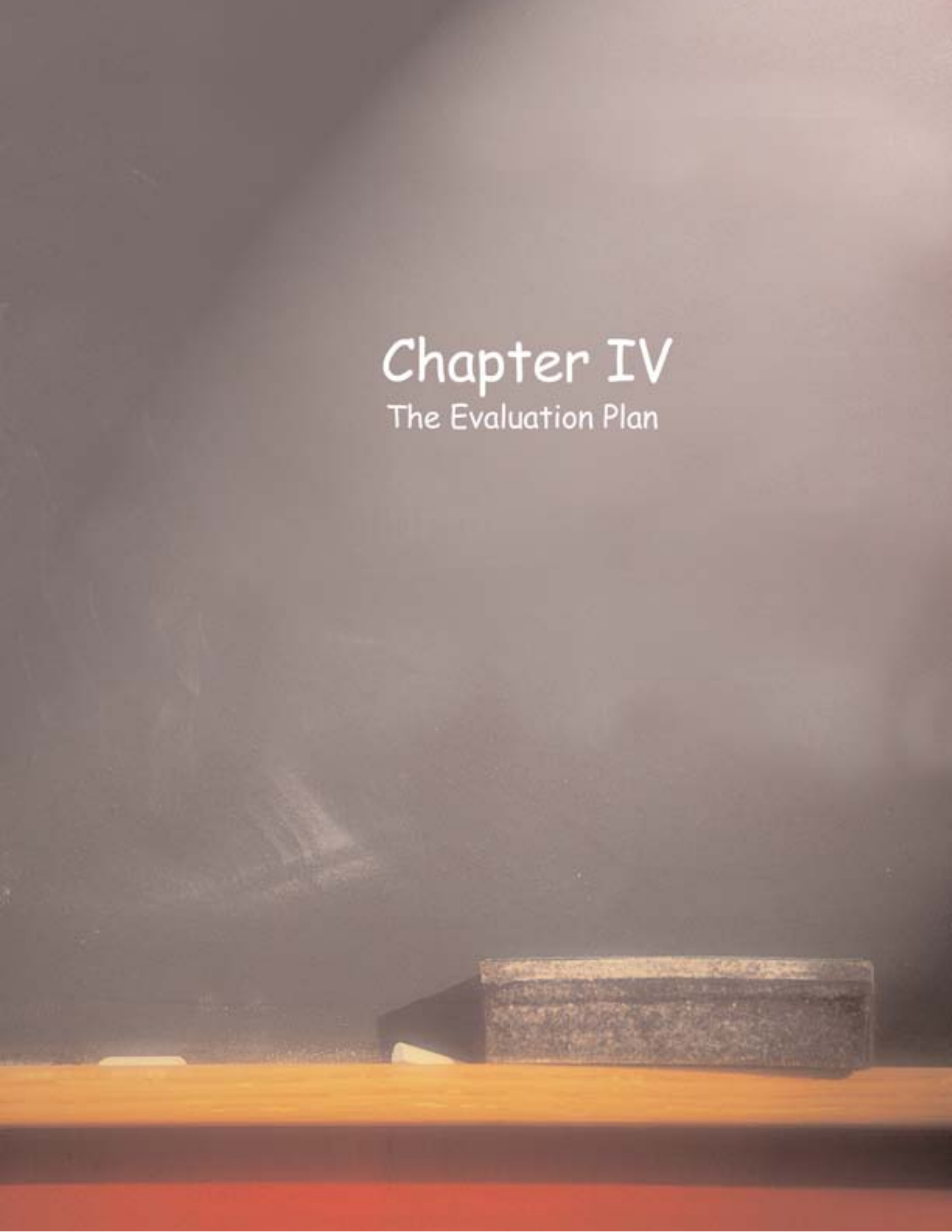
Implementing this critical feature requires deciding at what level operational decisions about resources will be made – e.g., within SLC, grade-level committees, school-wide committees, or school-wide committees consisting of representatives from each SLC or grade-level committee. Other decisions include determining who should be responsible for hiring new teachers, creating the school schedule, and choosing whether to purchase textbooks or an interdisciplinary curriculum.

While all schools were expected to implement activities associated with these reforms, the choice of specific activities targeting the seven critical features was left to each school. The FTF framework was not considered by its designers to be a "program" with a fixed set of materials and training procedures. Instead, it was intended to allow schools to work through a structured process for deciding how these seven changes were going to be implemented in their site.

Box E. What district-wide strategies are needed to build capacity and support school improvements?

Finally, to ensure that these change activities are implemented and sustained in schools, the district leadership (superintendent and district leaders, union leaders, and board of education) and other key community leaders are expected to create the conditions and build the capacity for change (Box E, Figure III-1). Creating the conditions for change requires ensuring that stakeholders in the schools and broader community understand the reform, are committed to the effort, are convinced that these changes can and will occur, believe support for the initiative exists and will continue, and believe meaningful change in student outcomes will be achieved. Building the capacity for change is the charge to the initiative's leaders to realign resources and provide supports as necessary to enable school sites to plan for, and initiate the critical features. These are the early outcomes that are considered necessary conditions for setting the stage for successful implementation of school-site improvements.

[10] In 2000-2001, IRRE changed this critical feature to "Equip, empower and expect all teaching staff to implement standards-based instruction that actively engages all students in learning."



Chapter IV

The Evaluation Plan

The FTF theory of change presents a road map to guide implementation of activities intended to produce systemic improvements. It specifies the goals – or outcomes – of the initiative, while the activities designed to reach these goals are chosen by the stakeholders working to implement the reform. Because the early, intermediate and long-term outcomes are specified in the framework, internal and external stakeholders can track progress, create accountability structures and make mid-course corrections in the reform process.

THE RESEARCH STUDIES

Several research studies were developed to test each component of the FTF framework as the initiative unfolded.

Initiative/District-Level Implementation Activities Study

A first step in understanding system-level change is to establish the historical context, motivators and investments involved in the change. This study documented how actions (planning and support) taken by the KCK district and its partners influenced the quality and depth of change in individual schools.

Early Outcomes Study

This study examined whether and how the Roundtables and Retreats held to introduce FTF to key stakeholders affected their attitudes and behaviors. According to the theory of change, these early outcomes are the necessary conditions for all stakeholders to sustain the level of activities required to initiate and maintain implementation of the critical features. The results of this study, along with the district-level activities during the early phase of the initiative, are contained in Gambone, Klem, Moore & Summers, 2002.

School-Level Implementation Study

The purpose of this study was to systematically study the **process and context** of planning and capacity-building activities at individual schools and their relationship to the **progress** of implementation of the seven critical features. Systems change requires significant planning, organizational restructuring, and capacity building in schools in concert with district-level supports, policy changes and re-allocation of resources.

Intermediate and Long-Term Outcomes Studies

The purpose of these two studies was to document (1) student and teacher experiences, and (2) student performance and commitment. These studies provide the necessary information to assess the association between changes in relationships, instruction, and school management structures and academic performance. These intermediate outcomes of supports and opportunities, and the long-term performance outcomes will be linked at the school-building level to both the earlier outcomes (quality and level of implementation) and the expected later outcomes (long-term development outcomes).

RESEARCH QUESTIONS

The studies described above were designed to address several research questions. Results for the District-Level Implementation Study and the Early Outcomes Study were described in an earlier report. This report presents results for the study of school-level implementation, and the intermediate and long-term outcomes studies. Guiding these studies were the following key research questions:

Question 1: Were the critical features of FTF significantly more likely to be present in buildings after three to four years of implementation?

Question 2: Did intermediate outcomes improve as buildings implemented FTF?

Question 3: Did long-term outcomes improve as buildings implemented FTF?

Question 4: What characteristics of leaders and staff in buildings seem to facilitate or impede implementation?

Question 5: Are improved classroom structures and instruction associated with better intermediate outcomes (student relationships with teachers and engagement in school)?

Question 6: Are improved intermediate outcomes (supportive relationships and engagement in school) associated with better long-term outcomes for students?

METHODS

Using a mix of surveys, teacher and principal interviews, classroom observations, and document review, the research team carried out a multi-method approach to data collection. Sources and timing of data collection for the research are summarized in Table IV-1.

Surveys were adapted from the Research Assessment Package for Schools (RAPS) (Institute for Research and Reform in Education, 1998). Surveys measured student and staff perceptions of the FTF reform initiative, as well as experiences of engagement and support. The surveys also served as a source of data for measuring implementation, as they provided a source of information on students' experiences in the classrooms and teachers' perceptions of how much change had occurred. Teacher and principal interviews focused on instructional change, school-level resource allocations and new management structures. Classroom observation protocols focused on documenting the frequency of specific instructional activities related to the critical features during core instructional time and were adapted with permission from the principal author (Stallings, 1977).

Basic information on the study measures is included in Appendix A. More detailed information, including copies of all instruments and a description of the analysis techniques used to construct the final measures used in the analysis, are included in the separate Technical Report.[11]

Analytic methods included the use of descriptive statistics and qualitative field note analysis (content analysis and theme identification). Inferential statistics – including tests of proportions and means, repeated measures analysis of variance, multiple regression and multi-level logistic regression – form the basis for conclusions about the validity of the theory of change and about changes from baseline (prior to the start of reform in each of the clusters of schools) in awareness, knowledge and commitment; levels of implementation; in opportunities and supports; changes in instruction; and differential levels of academic, social and long-term developmental outcomes.

ASSESSING CHANGE

Data were analyzed using a non-traditional method that shows results measured against a "standard" rather than the traditional method of looking at mean levels. This method shifts the focus from "group averages" to being able to distinguish between youth

Table IV-1: Timeline for Data Collection

Implementation Schedule						
Clusters	**1997-98**	**1998-99**	**1999-00**	**2000-01**	**2001-02**	**2002-03**
Wyandotte	Plan	Imp1	Imp2	Imp3	Imp4	Imp5
Washington		Plan	Imp1	Imp2	Imp3	Imp4
Harmon/Schlagle			Plan	Imp1	Imp2	Imp3

Imp indicates implementation

[11] This report is available online at www.ydsi.org.

Wyandotte and Washington Clusters												
	1997-98		1998-99		1999-00		2000-01		2001-02		2002-03	
Data Sources	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**
Staff Surveys		●		●		●		●*		●		
Student Surveys		●		●		●		●*		●		●**
Classroom Observation				●		●		●		●		
Classroom Observation Interviews				●		●		●		●		
Qualitative Data	●	●	●	●	●	●	●	●	●	●		
Stability Data					●	●	●	●	●	●	●	●
Records Data	●	●	●	●	●	●	●	●	●	●	●	●

** Survey revised.*
*** Includes engagement and teacher support only.*

Harmon and Schlagle Clusters												
	1997-98		1998-99		1999-00		2000-01		2001-02		2002-03	
Data Sources	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**
Staff Surveys		●		●		●		●*		●		
Student Surveys		●		●		●		●*		●		●**
Classroom Observation								●		●		
Classroom Observation Interviews								●		●		
Qualitative Data												
Stability Data									●	●	●	●
Records Data	●	●	●	●	●	●	●	●	●	●	●	●

** Survey revised.*
*** Includes engagement and teacher support only.*

who are doing well – defined as optimal levels – and those who are doing poorly – defined as risk levels. This method reflects the same type of approach now required by the No Child Left Behind Act (NCLB). It allows an assessment of the effectiveness of the reform in terms of moving students out of the bottom levels of outcomes (risk) and into the highest levels of outcomes (optimal).

Optimal and risk thresholds reflect critical levels for each measure (e.g., attendance rates above and below certain percentages). Students categorized as having a "risk" level of a particular outcome have a substantially increased risk for subsequent negative school performance compared with the baseline risk for all students. In contrast, students categorized as having an "optimal" level on an outcome have an increased probability of positive school performance compared with the baseline risk for all students. (See Appendix A for a description.)

Once data were coded into terms of "optimal" or high levels, intermediate or moderate levels, and "risk" or low levels, changes in the proportions of students and staff at each level were analyzed using multi-level, multi-variate logistic regression

techniques. Logistic regression techniques were employed because each of the outcomes – i.e., the threshold indicators (optimal or risk) – were coded as dichotomous variables.

The first step in the multi-level analysis is to construct models of the factors influencing change at the individual level. For example, in exploring changes in whether students experience support from their teachers, the first level model looks at how much can be explained by student demographic characteristics, how they perceive the standards and instruction in their classes, etc. Once as much variation as possible is explained with these factors, the next level model looks to see how much additional variation can be explained by factors "outside" the student – or at the building level. For example, continuing to explore differences in levels of teacher support, the building level model would then show – after taking out the variation from differences in individual students – how much of the difference among students can be explained by factors like characteristics of the staff, the student body composition, and how much implementation occurred in their building.

Technical descriptions of the analysis models and measures are presented in Appendix A. Full specification of the analysis models are presented in the separate Technical Report.

Because estimates of change are based on logistic models, the results are interpreted as the increased (or decreased) likelihood – or probability – that students will attain a particular threshold. For example, to estimate changes in reading test scores, two models were estimated: one for improvements (increases) in the likelihood that students would score at the proficient (i.e., optimal) level and one for improvements (decreases) in the likelihood that students would score at the unsatisfactory (i.e., risk) level. How to interpret these results is explained further in the next chapter.

ATTRIBUTING CHANGE TO THE REFORM ACTIVITIES

One of the key goals of the evaluation is to attempt to estimate how much of the change that occurred in KCKPS over the six years of the study can be attributed to activities associated with FTF reform efforts. For years, the "gold" standard in evaluation research has been random assignment studies, where control groups allow for estimations of the relative contribution of an intervention to any improvement observed in an outcome. However, as the field of social intervention has moved increasingly toward community-level solutions (and communities can rarely be "randomized"), a different approach has become necessary. Basically, a research project must be designed in a way that anticipates possible alternative explanations for change (meaning, that change is not due to the intervention) and include a strategy to test the likelihood that this alternative explanation is true (Granger, 1998).

One possible explanation for the improvements seen in the KCKPS District (shown in Chapter I) is that they were caused by changes in the student *population*. For example, if the overall ethnic composition changed in the district, or the SES composition of students changed over time, this could potentially explain increases or decreases in outcomes. In this study, every analysis included variables to remove the differences due to variations in demographics from our explanations of change.

Another alternative explanation for improvement in KCKPS is *historical*, i.e., changes were made at the federal and state levels (such as NCLB and state testing requirements) that affected all districts and schools regardless of whether they were implementing a reform. In this study, two strategies are used to account for this explanation. First, all analyses examining the change over time in an outcome are conducted using the number of years of implementation in the model rather than the calendar year. Because the reforms in the district were not implemented in all buildings in the same calendar year, we can examine the relative contribution of each additional year of implementation in a way that helps to minimize the effects of historical events that occurred in any particular calendar year.

Second, in what many would argue is the highest stake outcome (students' performance on state tests), we had the additional benefit of being able to compare our analyses of changes within the KCKPS District with those of the other districts in the state.

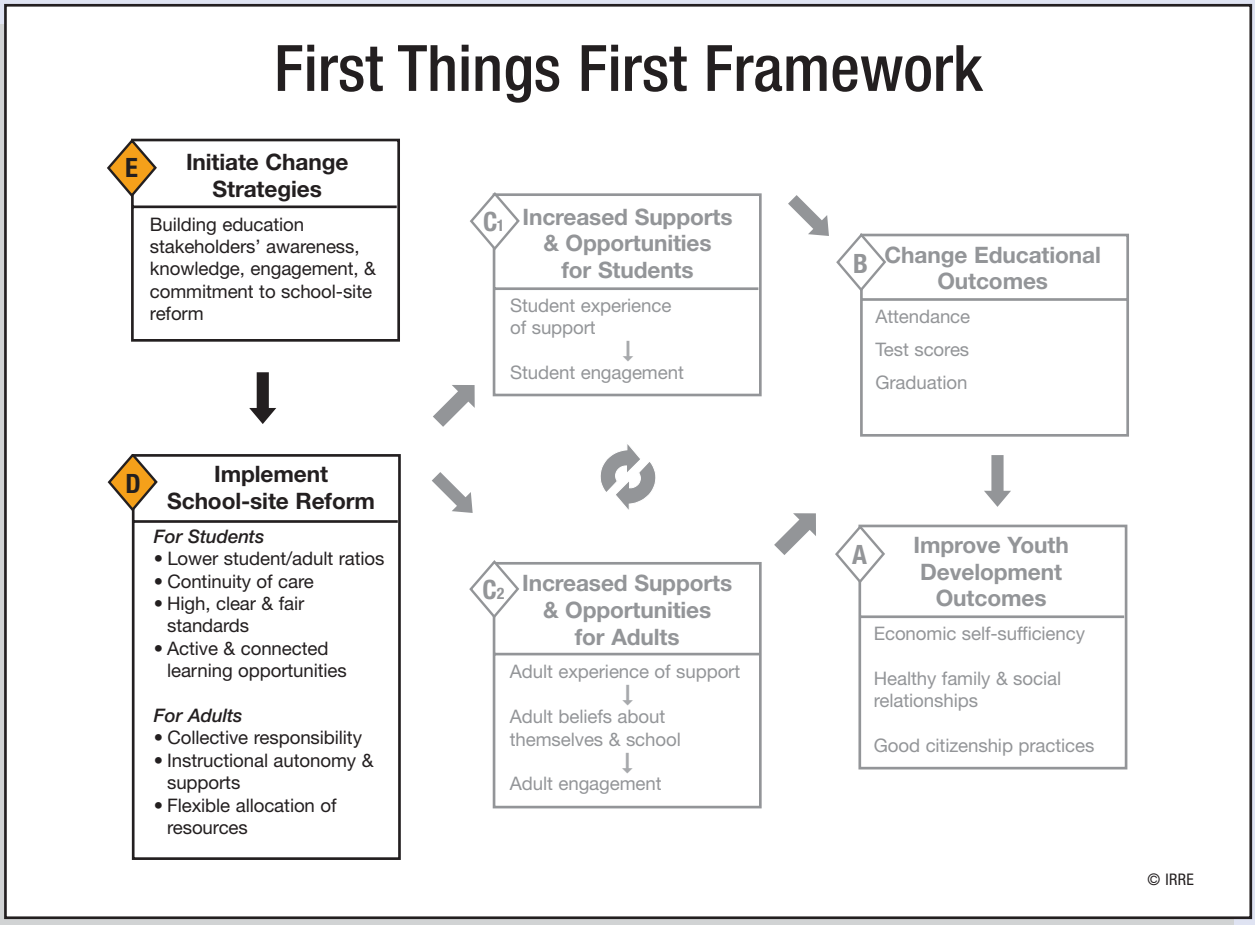
This, too, allows us to minimize the likelihood that the results seen in the evaluation are attributable to broader historical events rather than to any of the reform efforts being studied here.

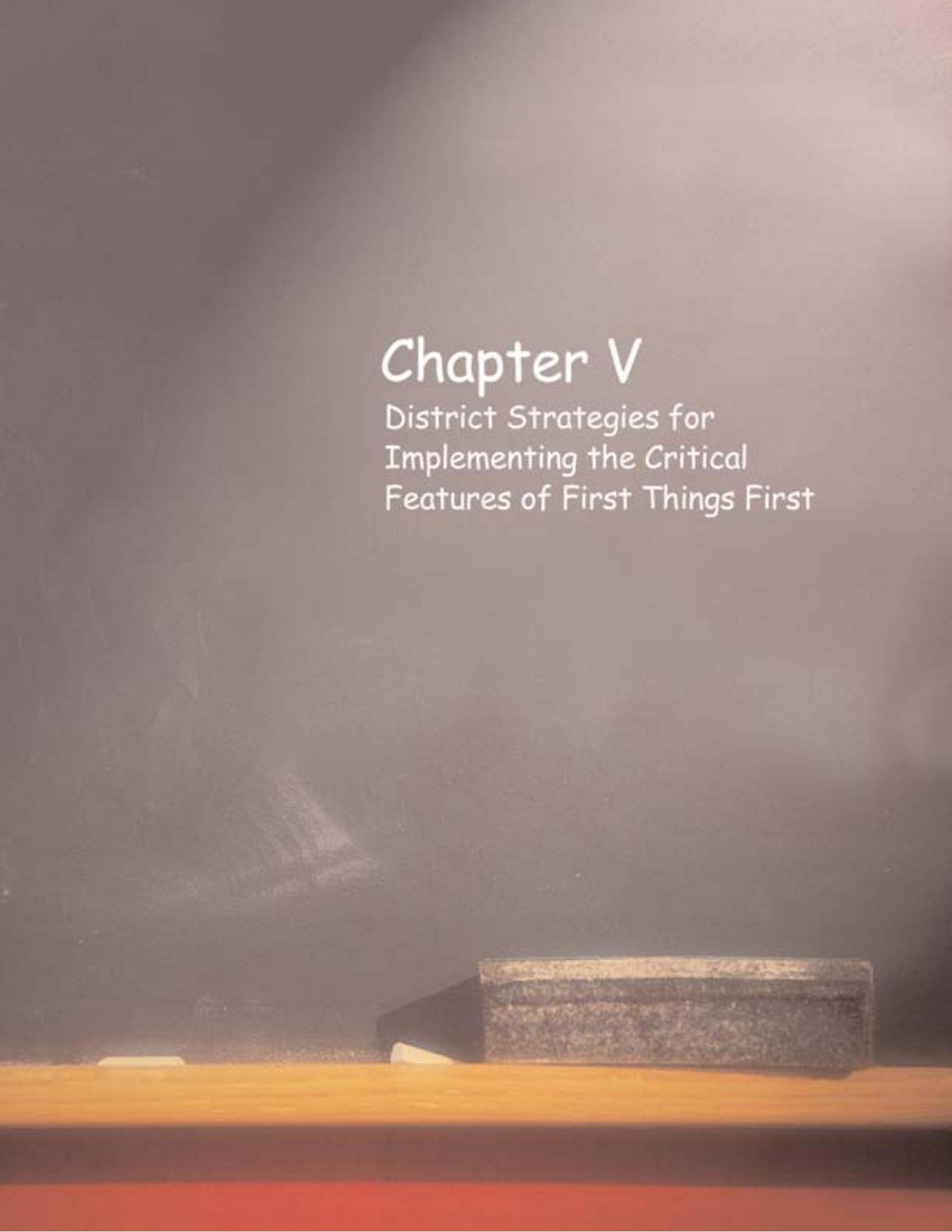
So, the design incorporates three sound strategies for assessing alternative explanations for change. But, because we were unable to also implement the intended matched comparison design, we are intentionally conservative about making any causal attributions in the analysis chapters. In the final chapter, however, we assess the evidence as a whole to consider the effect of FTF on outcomes and discuss the implications.

The remainder of the report presents the results of these analyses. The chapters are presented in the order of the reform framework, i.e., first we examine implementation, then intermediate outcomes and long-term outcomes. The last two analysis chapters examine the strength of relationships between each one of the elements of the framework as a final test of the theory of change. This is followed by the conclusion, discussing the implications of the results.

IMPLEMENTATION

QUESTION 1: Were the critical features of FTF significantly more likely to be present after three to five years of implementation?





Chapter V

District Strategies for Implementing the Critical Features of First Things First

At the district level, the central office was restructured to provide supports to the schools as they engaged in planning and implementing the reform. The restructuring involved creating the position of school improvement facilitators (now called instructional coaches) and giving them the responsibility of working with schools to plan and implement the structural and instructional parts of FTF. Further restructuring of the leadership itself involved assigning executive directors for each cluster and an exclusive focus on instruction rather than their previous broader responsibilities, which included operational oversight. (The conditions, strategies and resources affecting the District's efforts to build the capacity required to implement the reform are detailed in an early report on the initiative; see Gambone, Klem, Moore & Summers, 2002.)

In addition to structural changes, the District promulgated a series of policies designed to provide guidance in implementing FTF. These included a collective responsibility policy, a teaching and learning framework, guidelines for formation of Small Learning Communities (SLCs), and a set of curriculum standards and benchmarks. Further, the District began providing reports on student data that were organized by SLCs, schools and clusters. The purpose of these data reports were not only to sharpen the emphasis on collective responsibility but also to enable teams to design instructional strategies based on student performance.

These structural changes and policies were supported through resource investments (both through district funding and foundation support) and reallocations that enabled a more intensive focus on implementation of the critical features. For example, since 1998, the District has offered a two-hour "early release" time every Wednesday afternoon for the purpose of staff development. The continuity of this weekly focus on development has enabled extensive initiatives to enhance teaching and learning skills in literacy, math and general instructional procedures. A second significant investment has been the assignment of instructional coaches to every school; this investment was maintained in spite of state funding crises that have resulted in $13 million in cuts in the last two school years. Many other investments have been made, including an online curriculum library. The consistency of these investments was important not only in providing the resources needed for schools to implement the reform but also may have had a psychological effect of emphasizing the District's commitment to implementing FTF. For example, one district leader said:

> We eliminated one executive director in 2002-2003 and actually increased the number of instructional coaches. That passed a real strong message along because everyone believed we would eliminate that program [instructional coaches]…When we made the announcement [about keeping the instructional coaches], it represented a significant commitment for teaching and learning…This is no small decision.

Table V-1 provides a more detailed listing of actions taken at the district level to support implementation of each of the seven critical features.

UNDERSTANDING THE DATA AND RESULTS

Two types of data are presented in each of the remaining chapters: qualitative data that help describe and explain the reforms happening in the school buildings; and quantitative data that are the basis for making judgments about the extent to which implementation was successful and whether the chosen strategies contributed to improvements for youth.

The collection of qualitative data about implementation focused, by design, on the first two implementing clusters (Wyandotte and Washington) and was comprised of 14 elementary schools, four middle schools and two high schools. In the following chapters, descriptive information about how improvements were approached at the individual building level in these two clusters is based on written plans submitted by the schools, as well as interviews and observations by the research team. While there were many commonalities in approaches across the buildings, there were also some variations in approaches that reflected historical and contextual differences in these schools. Both the convergence and variations in building strategies are presented below.

The quantitative data, in almost all cases, are based on all four of the district's clusters of schools. The data on the last two clusters to implement FTF were originally not part of the evaluation, but rather were intended to be used for the district's management of the reform. This was because, in the original research design, these clusters would have only had

one year of implementation by the end of the study. However, an acceleration of implementation and extending the research for a year presented us with an opportunity to include these other clusters in analyses through three years of implementation. We chose to include the information here to strengthen the generalizability of the results and to take advantage of the additional analyses made possible with a larger sample of classrooms and buildings.

Each section in the remainder of this report follows a standard structure. The first page of each section has (1) a small chart to summarize where **statistically significant** (at .05 or better) improvements were made in outcomes after controlling for factors extraneous to FTF; and (2) a set of bullets describing the general findings for that section.

The implementation chapters then present information on the strategies that were used by the buildings in the Wyandotte and Washington clusters to put changes into place.

This is followed by a section that describes the **statistically significant increases** in the likelihood of positive outcomes and/or **decreases** in the likelihood

of negative outcomes over the period of the evaluation – with all clusters included for three implementation years. These changes are presented in line graphs that show trends in the changes in outcomes over the implementation period.

That is, they show the change in the likelihood of an outcome occurring, relative to the baseline year, for each additional year the reform had been implemented in a building. As noted above, all of the clusters

Table V-1: KCK Implementation of First Things First Critical Features

Critical Feature	Implementation Strategy
Provide continuity of care across the school day and across the school years	Divided schools into small teams of 5-10 teachers who teach the same group of 100-250 children for all four years of high school, all three years of middle school and all six years of elementary school.
	Implemented longer teaching periods each day. For example, at the high school level, periods went from an average of 45 minutes in length to 90 minutes.
	Provided a contact person at the school for every child and family, to be an advocate for that child and family. All staff at the school have a group of students for whom they advocate.
Lower student-adult ratios	Class size lowered in reading classes through redistribution of staff at each site.
	Instructional time in reading doubled from 50-100 minutes (average) every day.
	Additional periods of teaching time (30-60 minutes a day), in small groups (2-5 students per group) or with an individual student, are provided for students who are below grade level in reading and math.
Set high, fair and clear academic and behavioral standards	Trained and supported teachers in teaching to high standards: ● Literacy Academies ● Balanced Literacy, Just Read, Read Aloud and Read 180 Programs ● Teaching and Learning Coaches ● New math curriculum implemented
	Created easy access, through technology, to teaching materials that reflect high standards.
	Publicized to parents the district's high standards and shared with them ways they could help their children reach those high standards.

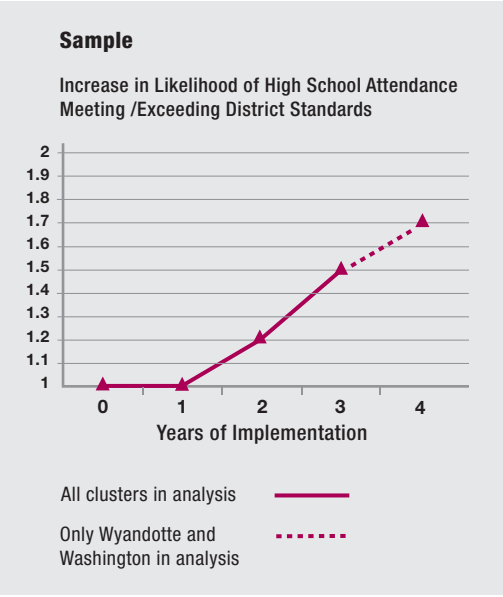
Table V-1: KCK Implementation of First Things First Critical Features	
Critical Feature	**Implementation Strategy**
Provide enriched and diverse opportunities to learn, perform and be recognized	Set district-wide goals requiring the standards to be taught and for all students to participate.
	Teaching and Learning Instructional Framework created that included nationally recognized teaching techniques to be implemented by teachers.
	All teacher training in the district is focused on teaching to high standards and ensuring participation by all students: ● Teaching and Learning Academies ● Kagan model for student engagement ● SIF/principal study groups ● Instructional coaching, with walk-throughs
Instructional autonomy and supports	Created teacher-training time every Wednesday afternoon for two hours.
	Created executive directors of instruction to focus only on instructional improvement.
	Assigned school improvement facilitators/instructional coaches to all schools (a total of 50). Over the course of the initiative, the central office transitioned from 1 SIF/2 schools in the Wyandotte and Washington clusters to only 1 SIF/2 schools district-wide. Then, some SIFs were reassigned as Literacy Coaches with a multiple school assignment. Finally, SIFs were assigned one per building.
	Invested in high-quality, nationally recognized teacher training for all teachers (see all items above under enriched and diverse opportunities).
Flexible allocation of resources (time, people, money, facilities	District reallocated Wednesday afternoon for staff development time.
	District shift of central office staff to schools as school improvement facilitators.
	District redirection of executive leadership positions from being responsible for all aspects of schools to an exclusive focus on instruction.
	School sites have increased authority to purchase teaching materials and hire personnel.
	Small teams of teachers and students have control over the resources impacting their students.
Ensure collective responsibility for student performance	Divided the district into four areas (clusters). Students stay in the same cluster for all 13 years of schooling.
	District distributes student performance data by: ● School and cluster ● SLC ● Family advocate group
	Developed a collective responsibility policy; all schools are mandated to develop and submit a collective responsibility plan, with goals for increased student achievement.
	Executive leadership of the district frequently visits schools to monitor the teaching of high standards and the participation level of students.
	High performance is recognized at all levels of the system through awards and public recognition.

are included in the data for the baseline through Year 3 of implementation.[12] However, the study period ended before the Harmon and Schlagle clusters completed a fourth implementation year. Thus, in select places where a Year 4 value is included on these charts, it represents only the buildings in the Wyandotte and Washington clusters. This is represented in the charts with a dotted line.

To interpret these "relative likelihood" results, it is important to note:

- When there is *no change* in the likelihood of a result between baseline and any year of implementation, it is represented by a *value of 1* in the graphs; therefore, all graphs (increases and decreases) start at the *value of 1*;
- Numbers greater than one represent an *increase* in the relative likelihood of seeing an outcome compared to baseline; and
- Numbers less than one represent a *decrease* in the likelihood for that outcome.

For example, suppose the outcome being examined is the attendance levels of high school students. Assume that, prior to FTF implementation, 50 percent of high school students were attending school regularly. If, after one year of implementation, the attendance rate remained at 50 percent of students coming to school regularly, the graph would show a value of "1" for that year. This would indicate there was no change. In this example, suppose after three years of implementation, the proportion of high school students attending school regularly increased to 75 percent. This would mean students were one and a half times more likely to have good attendance after three years of implementation than they were at baseline (75% at Year 3 divided by 50% at baseline equals 1.5). This is represented by the value of 1.5 on the line chart at Year 3 of implementation.



Another way to express such a result, which is also used in the report, is that for every 100 students who were attending school regularly at baseline (i.e., before FTF implementation), an additional 50 students (or a total of 150) would be attending regularly after three years of implementation.

It is important to understand that the changes shown in the charts represent the changes in the data after they are adjusted for the factors not related to FTF, i.e., changes in student demographics or changes in historical context. This allows us to more closely isolate the improvements associated with the reform. In many cases, were one to examine the unadjusted trends in the raw data in the district, they would not look the same as they do in these charts. The charts in Appendix B show both the unadjusted trends in the raw data and the trends adjusted to better isolate the effects of the FTF initiative. However, whenever the baseline value is important to understanding the estimates of improvements in an outcome, they are included in the body of the report.

Finally, results can also be expressed in terms of the percent change in the likelihood of an outcome. Continuing with this example, a change in the proportion of students with good attendance from 50 percent to 75 percent corresponds to a 50 percent increase in the likelihood of having good attendance.

Throughout the report, all trend graphs present changes in terms of the relative likelihood. The discussion of those results may include the alternative interpretations presented above to help clarify the meaning of the various results examined.

[12] This represents different calendar years in the different clusters.

A technical appendix – Appendix A – provides additional information about the measures and analyses. A separate technical report is available at www.ydsi.org with more detailed study designs, instruments, measure construction and analysis strategy.

EVOLUTION OF SIFS (SCHOOL IMPROVEMENT FACILITATORS) AS A STRATEGY FOR CHANGE

Initial Strategies:

- SIFs were hired from the curriculum coordinator staff and continue to be located in the central office. The initial SIFs were volunteers and report part time to the assistant superintendent for curriculum and part time to the director of school improvement, creating tension about workloads and competing priorities.
- Within three years, all curriculum coordinators become SIFs, even those who are resistant to the reform initiative. These SIFs – most of whom are assigned to the last two clusters to implement FTF – are less effective at helping staff through the planning process due to their own negativity. Ineffectual SIFs are eventually replaced with emerging leaders who choose to be SIFs.
- Some principals are threatened by SIFs and see them as an intrusion by the central office – SIFs deal with their relationships with principals in various ways, e.g., by making themselves "useful" to write various reports or working with staff who are struggling.
- The SIF's primary role is to work with principals and Stakeholder Committees to develop plans for restructuring the schools – with a focus on helping staff through their planning process.
- SIFs are assigned to two or more schools.
- SIFs meet together frequently and form their own Learning Community, which they use to develop skills for facilitation, instructional coaching and to clarify their roles – a group norm for inquiry and critical thinking develops.

Mid-Implementation Changes and Emerging Strategies:

- SIFs and principals both report to the same executive director who has responsibility for their cluster. Dual lines of reporting are eliminated.
- SIFs are "officed" in one of their two (or more schools) – some SIFs see this as a loss of a communication channel with the central office. The SIF's role is increasingly focused on professional development and instructional coaching. Most SIFs have major responsibility for structuring the Wednesday afternoon early release time.
- SIFs continue to meet together frequently, but also meet in teams with their schools (principal, literacy leaders) to work with the Literacy Academy.
- When resources for formal monthly Literacy Academy meetings are no longer available (Fall 2000), some SIFs are assigned to be Instructional Coaches and given responsibility for seven to eight schools.
- Most principals see their SIF as a critical resource; schools where the SIF is not officed believe they are getting fewer services from the SIF.

Current Strategies:

- SIFs and principals both report to the same executive director – two for elementary schools and one for secondary schools. One SIF and one principal from each cluster are designated as cluster leaders; this job is to be rotated.
- SIFs are retitled "Instructional Coaches" and there is a written job description developed by a committee of SIFs and administrators. The title change recognizes that the work of planning and restructuring is no longer a critical need, and that SIFs are evolving more toward instructional support and coaching. Instructional coaching responsibilities also fulfilled requirements to provide staff development specified by Title I and Title IIA; thus, the new title reflects the language of these programs.
- SIFs are assigned full time to one school, two at each high school.
- SIFs continue to meet together.
- SIF roles include both professional development during early release sessions, and hands-on, in-class observations and demonstrations – the focus is on engaging students and implementing the teaching and learning framework.

Lessons Learned:

- SIF roles need to be very clear, including relationships with all parties (especially principals).
- The SIF presence in the schools reinforces the idea that concepts learned in early release sessions are to be applied in the classrooms.
- SIFs need time to meet, to view themselves as continuously learning, which will enable them to transfer that learning to the schools.
- Investment in SIF positions provides a clear message to the schools that instruction is the first priority.

EXECUTIVE DIRECTORS AND CENTRAL OFFICE MANAGEMENT

Initial Strategies:

- The director of school improvement is responsible for supervising the planning of FTF. This position is changed to executive director of school improvement when the new superintendent is named, in order to give the position and the reform more credibility and power in the system.
- Two executive directors of operations are assigned two clusters each.
- Executive directors are responsible primarily for supervision of principals and focus heavily on budget, operations, etc.
- Instruction is primarily under supervision of the assistant superintendent for curriculum and instruction; the director of professional development also reports to this person.
- Some central office administrators do not believe FTF is relevant to their jobs – they may also see it as a reform that will come and go when the funding is over.

Mid-Implementation Changes and Emerging Strategies:

- The director of school improvement and one executive director of operations are appointed executive directors of instruction; the exclusive focus of the two EDIs is to be on instruction.
- Executive directors supervise SIFs and principals for their respective clusters, and also supervise professional development – placement of these development staff here produces some confusion with the roles of the Curriculum Standards Office.
- The assistant superintendent for curriculum is responsible for curriculum standards and benchmarks, and oversees curriculum specialists and the literacy initiative. Curriculum specialists (one for math, one for literacy, etc.) are organized as four cluster teams and assigned to clusters.
- One previous executive director is appointed executive director of instructional support. With the resignation of the assistant superintendent for business affairs, this person is appointed to that position.
- The superintendent emphasizes increased visibility in the schools and makes it clear that they, as well as the other administrators, are to attend Wednesday early release sessions, Literacy Academies and other events related to implementing FTF.
- Weekly management team meetings include all operational and instructional members; concern is expressed about communication "silos" (i.e., failure to communicate across different spheres of responsibility).

Current Strategies:

- After a brief period with four executive directors (one for each cluster), budget cuts require cutbacks; three executive directors are assigned: one for secondary schools, two for the elementary schools. Cluster teams are also dropped due to budget cuts.
- All administrative divisions (e.g., Human Resources, Technology Support) are expected to provide support to the schools for the instructional mission.
- Executive directors spend significant time in the schools; the presence of all administrators is more visible to staff throughout the district.
- The management team is expanded to include principals from each cluster, the ESL coordinator, etc.
- Executive directors, the superintendent and the assistant superintendent for curriculum meet as the instructional management team weekly, just before the larger management team – instructional decisions are made in this group. the instructional leadership team meets weekly on Wednesday; the focus is still the same.
- The assistant superintendent for curriculum is responsible for development of the online curriculum.

Lessons Learned:

- Development and change must be viewed as everyone's responsibility and not just one person's (i.e., there should be no "director of instructional improvement").
- The lead central office figure charged with the reform must have positional power within the system. The move of director to executive director of school improvement signaled the importance of the reform work.
- Reporting and communication lines need to be clear and not "crossed."
- Organizational structure should facilitate a focus on instruction as the primary focus of administrators.
- Administrators need to view themselves as supporting instruction rather than monitoring compliance in the schools.
- Continuous improvement and development must be seen as "the way we do business" rather than a (by implication, temporary) reform.
- Administrators need to be visible role models, demonstrating a willingness to change at the central office, to provide supports to the schools, and to participate in the professional development occurring at the building level.
- Messages about the focus on instruction need to be consistent and backed up with resources and support.

PROFESSIONAL DEVELOPMENT: EARLY RELEASE, TEACHING AND LEARNING

Initial Strategies:

- Kauffman initially funds extra in-service days to be taken in the summer just prior to implementation of the school's FTF plan; administrators quickly realize this is insufficient.
- District-wide Wednesday afternoon early release in-service sessions are first implemented in Year 2 of the project (first implementation year for Wyandotte, planning year for Washington).
- The Board of Education works with the community to address issues and concerns of parents regarding Wednesday afternoons.
- Central office administrators work with the teacher's unions and the community to allay concerns about the early release; the superintendent makes numerous presentations to community groups throughout Kansas City and receives commitments from religious and social service groups to provide services to students on Wednesday afternoons.
- Parents and other community members are invited to attend in-service sessions to see that teachers are working and to understand the purpose.
- The initial format for the in-service days involves learning to disaggregate student data – the plan is to develop instructional strategies around student needs for each school.
- Central office planning for the in-service sessions is barely a week ahead of presentations in the schools; the schools quickly get ahead of planning; there is little support for the idea of approaching development with student data analyses.
- The decision is made in the spring of Year 2 to develop a literacy initiative; key teacher leaders are identified to participate in Literacy Academies and to provide support to staff to develop literacy skills in all schools, K-12.
- The focus on literacy leads to a variety of reading program strategies, including Balanced Literacy (which involves using leveled groups), Just Read, Read 180, etc. Advocates for these different approaches are at odds with each other.

Mid-Implementation Changes and Emerging Strategies:

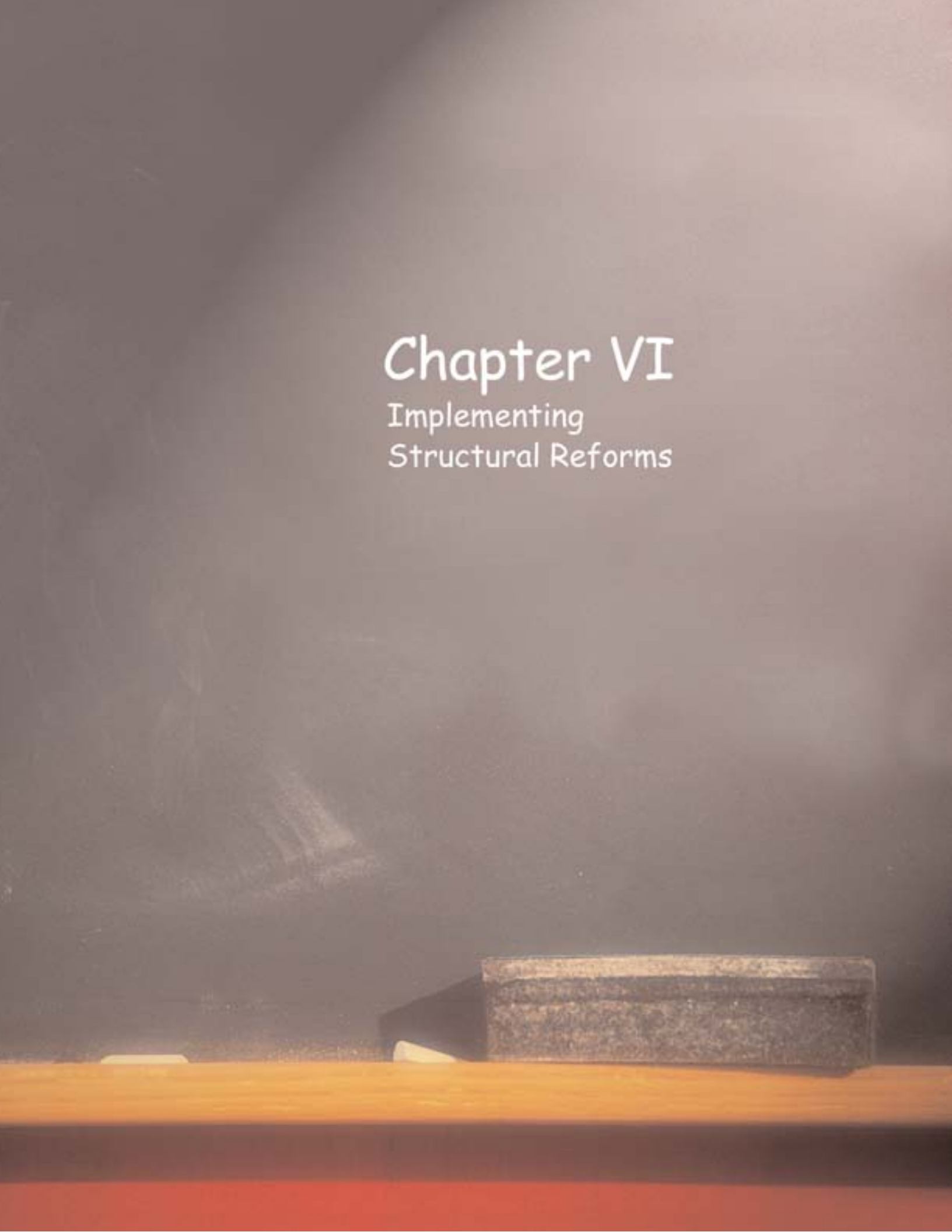
- The Literacy Academies, held monthly, provide the content for bi-weekly early release sessions, taught by the literacy leader, principal and SIF.
- The superintendent intervenes, making it clear that all literacy strategies are valuable, "silos" are unacceptable, and all components of the literacy initiative should work together.
- After the monthly Literacy Academies are scaled back due to resource shortages, the Instructional Coaches bring the Literacy Initiative to the schools. Later, this becomes the teaching and learning framework, based on two workgroups. The first teaching and learning workgroup is made up of SIFs and teachers, and evolves into a larger workgroup consisting of district and building administrators, teachers and union leadership. Union involvement in this process is a key move because it actively involved the union in the reform work. The process moves the union involvement beyond contract negotiations, with a tacit agreement to (at best) not get in the way of reform, but to integrally involve the union leadership in the focus on teaching and learning.
- On the "off" weeks, schools are free to focus on other topics identified by staff as professional development needs.
- In some schools, the first half of the early release session is spent in whole-group time and the second half in meetings of SLC teams.
- There is little continuity from week to week to follow up on whether skills taught in early release were implemented in classrooms.
- The Board of Education continues to require strong justification from the district administrators to approve continuation of the Wednesday early release times.

Current Strategies:

- Evidence of improved student outcomes leads to widespread community acceptance of Wednesday early release sessions as an ongoing part of a continuous learning community
- Instructional Coaches (previously SIFs) and principals, along with teacher leaders (usually SLC coordinators), plan and present in-service sessions based on the teaching and learning framework.
- Administrators, Instructional Coaches and principals reinforce strategies presented in in-services through in-class "walk-throughs" and observations.
- Focus is on engaging students.

Lessons Learned:

- Community acceptance needs to be cultivated through providing a sense of urgency, openness and community-wide participation in the effort.
- Sessions need to be well-planned and structured.
- Involvement and active participation by the union in designing the framework for teaching and learning creates a sense of ownership and "buy in."
- Using group processes in learning and/or implementing some instructional strategies (e.g., read alouds) can help create a sense of purpose for SLC teams.
- Facilitators (SIFs, principals, literacy leaders, SLC coordinators) need training on how to facilitate the in-service sessions.
- There needs to be continuity in in-service content from week to week and during the week to reinforce in-class application.
- The weekly in-service sessions can help to create an organizational climate that encourages continuous learning and change.



Chapter VI

Implementing Structural Reforms

MAJOR FINDINGS

- *Structural changes were focused on (1) lowering student- adult ratios to 15 to 1 during reading and math instruction; and (2) creating more continuity of care for students by ensuring they stay in a Small Learning Community (SLC) for two or more years (i.e., SLC stability).*

- *In secondary schools in the Wyandotte and Washington clusters, there was a significant decrease in the likelihood of students being instructed in classes with student-teacher ratios that exceeded district standards (23:1). However, there were no significant improvements in student-teacher ratios in elementary schools.*

- *Students at all levels in the first two implementing clusters were more likely to experience the continuity of being assigned to the same SLC for two or more years as implementation of FTF occurred. At the elementary level, students were twice as likely to experience this type of continuity and at the secondary level this continuity increased fourfold after three years of reform.*

Improvement in Structural Implementation Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Observed Ratios				✔
SLC Stability	✔	NA	✔	NA

✔ ***indicates statistically significant improvement in outcomes***
NA ***indicates not applicable***

Ratios:
Optimal = LE 15:1
Risk = GE 21:1 elementary; 23:1 secondary

Stability:
Optimal = students in SLC two or more years

HOW WERE STRUCTURAL REFORMS IMPLEMENTED?

Two goals of the FTF initiative were to reduce ratios and increase continuity of care. The primary mechanisms for this were structural changes in schools.

Lower Student-Adult Ratios

Building staff had to decide how to structure the school to achieve a 15 to 1 student-adult ratio during the core instructional periods (i.e., reading and math).

Elementary Schools

All 14 elementary schools in the two clusters proposed using "special" teachers (physical education, music and art teachers, special education staff), support staff, and aides or paraprofessionals during a designated core instructional time for reading and math. The Wyandotte cluster approaches generally entailed adding these adults into the classrooms to lower ratios during core instruction. In contrast, schools in the Washington cluster added these additional adults to the classrooms, but also used different strategies for grouping students during reading and math instruction (e.g., leveled reading groups[13] or multi-aged groups[14]). The difference in the approaches may be due to the (a) more specific guidelines introduced by the district in 1998, and (b) introduction in 1998 of the literacy initiative, with its emphasis on leveled instruction, and the early release professional development times. Both of these developments occurred after the plans for the Wyandotte cluster had been written.

Secondary Schools

In the six secondary schools, there were two primary strategies for achieving lower student-adult ratios. Two Wyandotte schools used block scheduling of core instruction within SLCs. The remaining schools implemented a "power hour" or skill-builders class, taught by all staff during block times.[15] A variety of non-core staff, volunteers and other staff supplemented core instructional staff during these block times. All six secondary schools also proposed the use of special subject teachers (e.g., art, music, physical education), non-core instructional staff, administrators and community volunteers during core instructional time. The implementation of the literacy initiative, including the use of leveled instruction, also influenced the organization and some variation in size of some of the power hour groups, with smaller groupings of students at lower reading levels.

Implementation Influences

Shifts – either increases or decreases – in enrollment created challenges to schools that chose to use redistributions of adults into classrooms during core instructional time. Similarly, these enrollment shifts might also result in changes in available funds (thus, gains or losses in paraprofessionals) through Title I or special education allocations. This was especially true of smaller elementary schools where enrollment declines caused them to share their "special" teachers with another school or eliminated these teachers from the core instructional time altogether, which caused scheduling difficulties. At the secondary level, the power hour/skill-builder strategy appeared to be more impervious to shifts in enrollment.

Continuity of Care

Schools had to decide how to reorganize to allow smaller groups of students to have contact with fewer adults for longer periods each day and for more than one year. Two basic strategies were used to implement continuity of care: the SLCs and looping.

[13] Leveled reading instruction involves matching the content of the work (e.g., reading text) to the child's instructional level. This type of instruction is believed to be the most effective form of literacy instruction.

[14] Multi-aged groups place students of the same instructional level together regardless of age. In some cases, multi-aged instructional groups will place higher- and lower-skilled students together so that they can learn from each other.

[15] Power hour is a reading and math intervention, where every two weeks, a student has five extra instructional hours in addition to their normal daily reading and math instruction. The interventions are leveled so that ability and instructional content are matched. Student-adult ratios are often 10:1 or less.

Elementary Schools

All but one of the elementary schools initially proposed some form of SLC, referred to in their plans as "houses" or "families." Of those schools with SLCs, all but one organized their SLCs roughly with at least one classroom for kindergarten through fifth grade. All but three of the elementary schools chose some form of looping. One school was organized in a multi-age grouping strategy, while a second used a multi-age grouping for only one of its SLCs. Finally, two schools did not propose looping, but rather engaged in team teaching with teachers in the grade levels above/below them. By 2001, however, all schools were mandated to organize in vertical SLCs that included at least one classroom per grade level. By the end of the study, most looping strategies included at least two years of looping in K-2 and 3-5 (primary and intermediate levels).

Secondary Schools

Students were assigned to SLCs in all secondary schools. Initially, two basic patterns of design in SLCs included (a) vertical SLCs involving all the middle or high school grades, with looping; or (b) "upper" and "lower" division SLCs (e.g., grades 9-10 and 11-12). SLC organization was based either on themes (e.g., health careers, arts) or on ability or other special groupings (e.g., all English as a Second Language students). By Spring 2001, however, the leadership team decided that to be effective, all secondary schools should be required to develop plans for vertical SLCs and to do away with any kind of "tracking" or ability grouping in SLCs.

Implementation Influences

At the secondary level, the design and implementation of the continuity of care feature had become relatively standard by the end of the study period. At the elementary level, several factors led to variations. There was tension between the vertical SLC concept and teachers' perceived need to meet and plan together at the grade level. Most schools compromised by alternating grade level and SLC or vertical team meetings. Also, several elementary schools implemented "Fifth Grade Academies," based on the rationale that fifth graders needed a specialized preparation for transition to middle school, which pulled fifth graders out of their SLCs into this new configuration. In still other elementary schools, the enrollment patterns were such that it was not always possible to have every grade in every SLC; for example, one SLC at a Washington cluster elementary school did not have a fourth grade.

HOW MUCH WAS ACHIEVED IN REFORMING STRUCTURE?

In much of the report, where student survey and records data are used, we examine the effects of the reform on all four clusters in the district. However, the results presented here on the progress made in reforming the structure for instruction in buildings (classroom ratios and SLC continuity) are limited to the Wyandotte and Washington clusters. This strategy enables us to examine three years of implementation in both clusters. If we had included the Harmon and Schlagle clusters, we would only have been able to examine improvements after two years of implementation since there are no Year 3 observation data for these clusters.

We collected data on classroom ratios during the structured observations conducted every year. Each of the approximately 250 classrooms in Wyandotte and Washington, where observation data were collected each year, were coded for the student-teacher ratio. The goal of implementation was to reach a ratio of 15 to 1. Ratios of more than 21 to 1 at the elementary level or 23 to 1 at the secondary level were outside the standards set by the district.

Data on the continuity students experienced in their SLCs come from the student records data maintained by the district – each year, buildings recorded students SLC assignments. Here, we examine the proportion of students in a building who were in a SLC for at least two years (representing continuity) versus the proportion in a SLC for less than two years. Since students could not be in SLCs for two years until implementation was underway for that period, we consider the second year of implementation the baseline for this measure.

RATIOS[16]

The observation data show that in secondary schools there was a significant improvement in the likelihood of students being in classrooms that exceeded the district standard (23:1). Chart VI-2 shows that for every 100 students in secondary schools that were in classrooms with higher than desired student-adult

[16] Charts with the relative likelihood of being in the optimal and risk categories for each variable after three years of implementation are included in Appendix C.

Chart VI-1

Increase in Likelihood of Students Instructed in Classrooms with Low Student-Adult Ratios

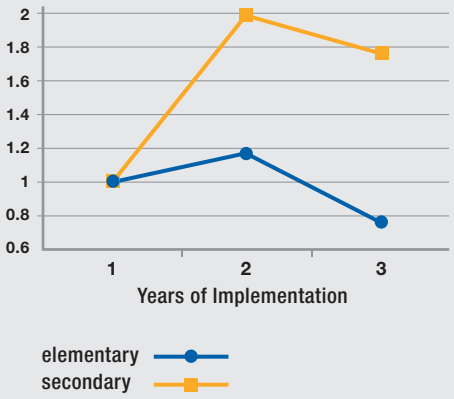
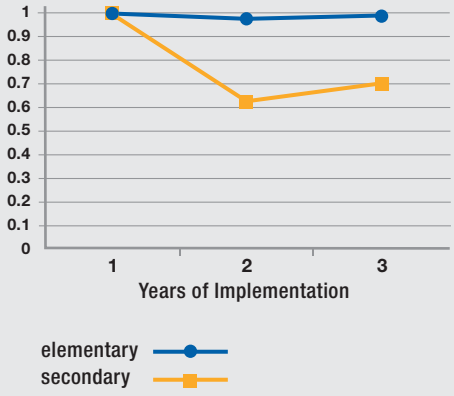


Chart VI-2

Decrease in Likelihood of Students Instructed in Classrooms with High Student-Adult Ratios



ratios in the first year of implementation (represented by the value 1 on the chart), only 70 students were being instructed in classes of that size after three years of reform in the Wyandotte and Washington clusters (represented by the value of .7 on the chart for Year 3 of implementation). We did not see a significant improvement during this period in the likelihood of students being in the optimal-sized classrooms with the FTF desired ratio of 15 to 1.

For elementary schools, the data show a significant result in the wrong direction for the likelihood of students being in classrooms with the desired 15 to 1 ratio after three years of implementation in both clusters (see Chart VI-1). It appears that within the first three years of implementation observed, student-adult ratios worsened in elementary schools; while in the secondary schools, there was a significant decrease in the proportion of students in classrooms that were too large. A review of significant events occurring in the district suggests some possible explanations:

- At the elementary level, two events coincided that may explain the loss of progress in student-adult ratios. First, the district experienced an acute teacher shortage that peaked in the 1999-2000 and 2000-2001 school years; this shortage was caused, in part, by a number of teachers retiring (a national trend with many "baby boomers" beginning to retire) and by an inability of the district to compete with salaries offered in nearby suburban districts. By 2001-2002, this trend had begun to reverse, as some of the District's/Kauffman's initiatives (Grow your Own, Teaching Fellows) began to have an effect on the number of teachers.

- Second, the 10-year decline in enrollment in the district came to a halt beginning in Fall 2000, when enrollment stabilized. The stabilization was due, in part, to an influx of Latino families into Kansas City, Kansas. The larger number of students enrolling in some schools, coupled with the teacher shortage, may be a partial explanation for the step-back in progress in reducing student-adult ratios during core instructional time.

Small Learning Community Stability

One of the major variations in implementation of SLCs that occurred in KCK was the evolution of the requirement for schools to implement the SLC structure with multiple grade levels for continuity. For the first implementing cluster (Wyandotte), this was not a requirement until their second year of implementation. For the second implementing cluster (Washington), a decision was made during their planning year to require SLC structures in the plans. This change encountered some early resistance from staff in some Washington buildings who believed they were not being given the same decision-making autonomy that Wyandotte cluster schools were given. This dynamic, coupled with the fact that a fair amount of "shaking out" occurred in learning how to best implement this feature, resulted in SLC structures that changed in some buildings over the first few years of implementation.

When we looked to see if there were improvements in the likelihood of students being in SLCs for two or more years, we found it doubled for elementary students and increased fourfold for secondary students (see Chart VI-3). This meant that by the third year of implementation, an additional 100 elementary students and 300 secondary students were in SLCs for more than two years for every 100 students who had this type of continuity after only two years of implementation. The trends illustrated in Charts VI-3 and VI-4 show that this improvement persisted in the fourth year of implementation in the Wyandotte cluster.

Chart VI-3

Increase in Likelihood of Students Experiencing High SLC Stability

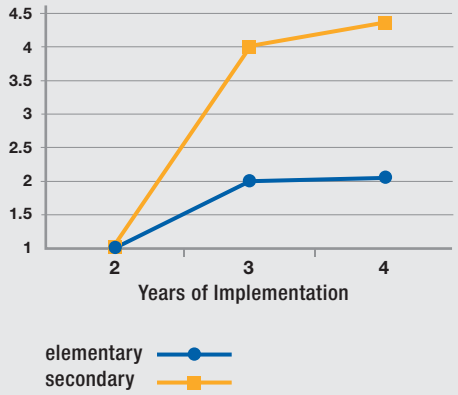
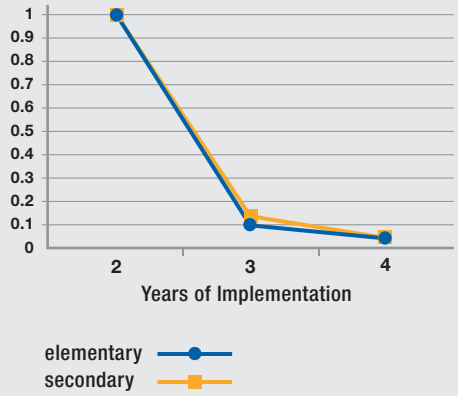


Chart VI-4

Decrease in Likelihood of Students Experiencing Low SLC Stability



ELEMENTARY SCHOOL EXAMPLE

Structural Implementation Outcomes

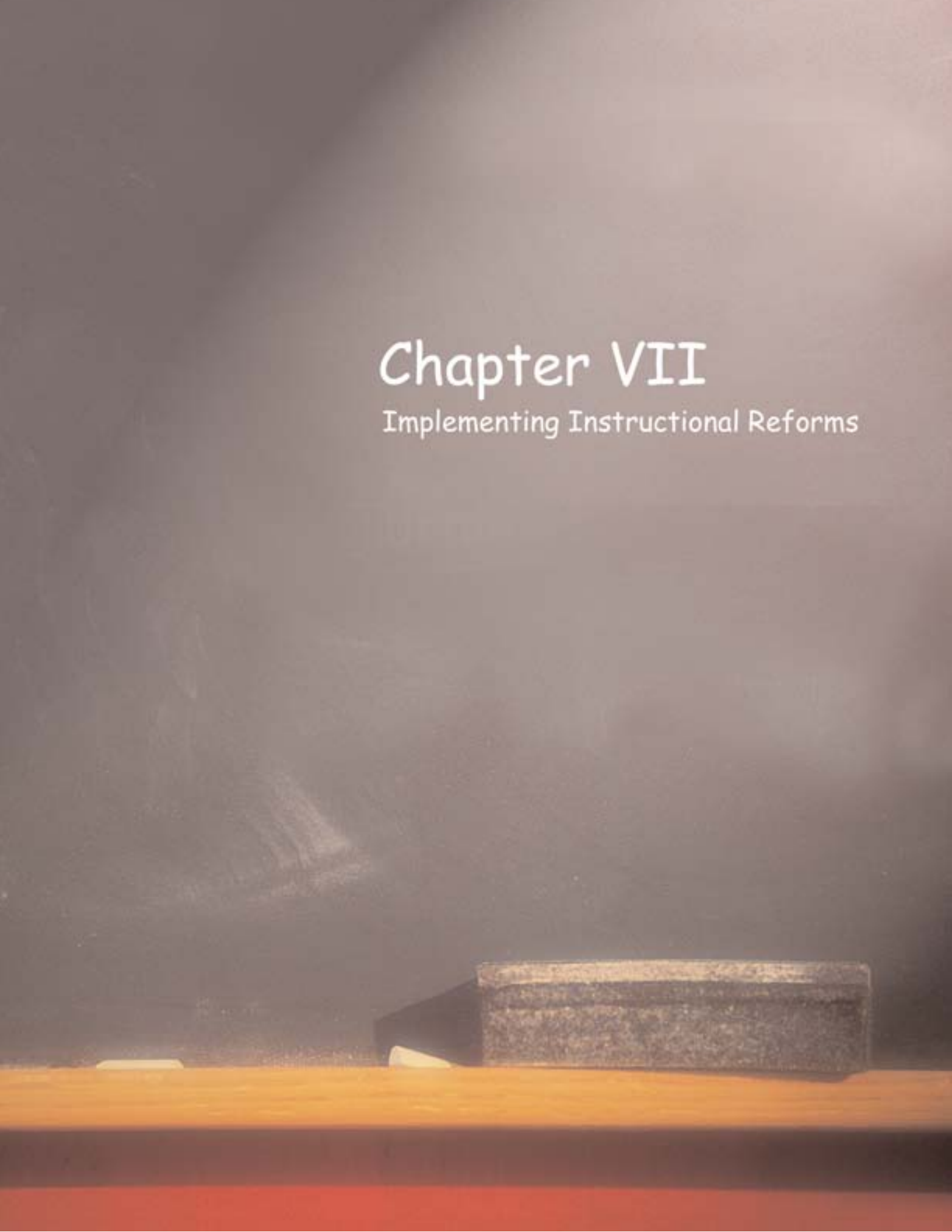
Student-Adult Ratios: A Wyandotte elementary school provides an example of how changes co-occurring in the district may have affected the observed student-adult ratios. The FTF initiative coincided with the final implementation of the district's desegregation plan. Catchment boundaries were redrawn just prior to implementation of FTF, resulting in shifts in both enrollment numbers and demographic patterns. Connected to the boundary shifts, three schools were slated to be closed while a new building was constructed. During the years of closing, construction and re-opening, some students from schools slated for demolition were relocated temporarily to another, with others distributed to two other elementary schools. With the completion of the new school building, these students were brought together again in one school. The result was that enrollments shifted from 199 (with 13 staff) in 1999-2000 (ratio = 15:1) to 633 (38 staff) in 2000-01 (ratio = 17:1).

A different elementary school in the Washington cluster provides an example of how even small shifts in enrollment, especially when it is under-projected, can upset plans for achieving the lower student-adult ratios. In this case, the "special" professionals (e.g., physical education, music, art, special education) were assigned to two or more small schools. The fact that these staff members were only in the building part time resulted in reduced flexibility and reduced available adults during core instruction time. In the planning year (1998-1999), the staff at the school planned to lower student-adult ratios by having primary and intermediate grades in reading and math for a two-hour block each day; utilizing classroom, special education, KALL (a specialized reading development program for primary grade teachers), Title I teachers and special education paraprofessionals, and Title I aides. Because of increases in student enrollment in 1990-2000, intermediate grades experienced difficulty in maintaining a lower student-adult ratio. The primary grades had lower numbers and experienced more success in program implementation. In 2000-2001, the student-adult ratio increased from 15:1 to 18:1 and, in some cases, 21:1 during the reading block time, because of shortages in special personnel and an under-projected student enrollment. Enrollment was underestimated by 28 students. Schedules were altered and the school had difficulty implementing protected blocks in reading and math. The principal said that they did the blocks when they could, and that if she could add extra bodies it would lower the class size: "I can get two teachers for one KALL teacher. This program is draining the teacher pool. We have excellent teachers who are learning to be KALL teachers. I am not against the program if they want to hire teachers from outside the district." The school added one more classroom in 2001-2002 to reduce the student-teacher ratio from 29-30 students to 23-25 students. However, the lack of "specials" who were dedicated to this school exclusively reduced the ability of the staff to include these teachers in the core instructional times of most classes, and resulted in a failure to achieve significant further reductions in the student-adult ratio.

SECONDARY SCHOOL EXAMPLE

Structural Implementation Outcomes

Student-Adult Ratios: A typical strategy for implementing lower student-adult ratios during core instructional time in the middle and high schools was the implementation of a "power hour" or a "skill-builders" class held at the beginning of the day. For students with special needs or who are otherwise "at risk," these basic or core instructional classes tended to be smaller and more individualized. We learned that these "power hour" strategies were a useful tool for achieving the lower student-adult ratios; however, we also learned that the lower ratios did not always result in improved student engagement and instruction. For example, one researcher who "shadowed" a special education student as he went through his day at a Washington cluster middle school noted the following: In skill-builders class, which is first thing in the morning from 7:45am to 8:45am, there were only 11 students. The students were doing math worksheets. Half of the students in the class were not engaged, combing their hair and visiting. Two of the students were sleeping until they were handed a calculator to use during the last part of the period. After skill-builders class, the student went to reading class. Again, there were only 11 students in the reading class, which was using Read 180. The student being shadowed and about half of the students in the class worked on the computer with the Read 180 program, while the other half of the students in the class worked with the students in a round-robin reading activity. All of the students were actively engaged in their work. Regardless of class size, teachers still need to actively engage their students in the classroom.



Chapter VII

Implementing Instructional Reforms

MAJOR FINDINGS

- *After the initiative leaders sorted out the preferred approach to structuring schools, they turned their attention to improving instruction. There were significant improvements in the grouping formats used for instructing students at all levels, the types of teaching/learning strategies used in high school, and the consistency of high academic and behavioral standards at all levels.*
- *In both elementary and secondary schools, the use of small grouping strategies during instruction was observed in more than twice as many classrooms after implementation of FTF.*
- *There was no clear and steady trend in the likelihood of observing active instructional strategies being used by teachers in classrooms over three years of implementation of FTF. However, students in high schools were significantly more likely to report having opportunities to work in teams on assignments, and to work on projects that connect subjects and are connected to their futures and lives outside of school.*
- *Students were significantly (up to four times) more likely to experience classrooms with high academic and behavioral standards during the course of FTF implementation.*

Improvement in Instructional Implementation Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Observation of Small Grouping	✔	✔	✔	✔
Observation of Active Learning			✔	✔
Student Report of Instruction			✔ (H)	
Student Report of Standards	✔	✔	✔	✔

✔ ***indicates statistically significant improvement in outcomes***

Observation of Small Grouping:
Optimal = students spend 2/3 or more of observed time in small groups
Risk = students spend 1/3 or less of observed time in small groups

Observation of Active Learning:
Optimal = students spend 60% or more of observed time in active learning
Risk = students spend less than 25% of observed time in active learning

Student Report of Instruction:
Optimal = at least 9 of 12 items answered "sort of true" or "very true"
Risk = at least 4 of 12 items answered "not very true" or "not at all true"

Student Report of Standards:
Optimal = at least 7 of 10 items answered "most of the time" or "almost always"
Risk = at least 6 of 10 items answered "not very often" or "almost never"

HOW WERE INSTRUCTIONAL REFORMS IMPLEMENTED?

Enriched and Diverse Learning Opportunities

The understanding of the meaning of this feature evolved over time, from the initial belief that it referred to such things as field trips, project-based learning, etc., to an explicit teaching and learning framework focused on developing instructional skills aimed at engaging students in higher-order thinking.

Elementary Schools

All but 3 of the 14 elementary schools in the Wyandotte and Washington clusters initially approached this feature, in part, by planning a variety of learning activities, special programs or other initiatives in their school designed to enhance students' learning experiences. Beyond that commonality, two interesting patterns emerged in the plans for this feature. First, all but one of the Wyandotte cluster schools selected the dimensions of learning framework as the basis for their initial approach to this critical feature, while only one Washington cluster school did so. A possible explanation for this difference may be that the dimensions of learning framework was espoused by an instructional committee under the leadership of the former associate superintendent; several of the participants in that district-level committee later became Wyandotte cluster SIFs. Second, more of the Washington cluster schools plans (the second implementation cluster) moved to learning approaches such as Balanced Literacy, multiple intelligence approaches[17] and recognition of student achievements (e.g., publicly acknowledging increases in student performance). As time passed, the teaching and learning framework took the place of the dimensions of learning framework,[18] allowing such strategies as the Balanced Literacy program, the Read Aloud program and the Six-Trait Writing approach to be incorporated. In Fall 2001, individual elementary teachers began to pilot the new math curriculum – Math Investigation. In Fall 2002, all elementary schools were implementing units from Math Investigation.

Secondary Schools

All but one of the six secondary school plans mentioned staff professional development as one of their proposed tools for enriching learning opportunities. In addition, three interpreted this feature to mean more activities and experiences were needed for students (e.g., portfolios, technology, interaction with the community). As in the elementary schools, secondary professional development focused on the teaching and learning framework, and included strategies for student engagement, the Kagan cooperative learning methods and the read alouds. In Fall 2001, the new secondary math curricula (Connected Math for middle schools and Integrated Math Project for high schools) was piloted by several teachers. By 2002, secondary schools were implementing many of the units from these math curricula.

Implementation Influences

After the initiation of the Wednesday afternoon early release sessions, implementation of this feature varied across the years as much or more than it did within different schools. The early release time sessions were initially focused on helping staff disaggregate student data in order to base curriculum objectives on student performance. This approach evolved into the Literacy Academies, with an initial focus on reading instruction. The monthly Literacy Academies were attended by principals, SIFs and identified literacy specialists from each school. Substitute teacher shortages in Fall 2000 led to the creation of instructional coaches, who delivered professional development around the literacy initiative at the building level. At the elementary schools, the Reading in Excellence grant was another vehicle for delivering literacy instruction; this was an intensive mentoring development program that focused on Balanced Literacy. Toward the end of the study

[17] Multiple intelligence approaches to instruction take into consideration the idea that every individual has different kinds of abilities through which he or she learns in addition to intellectual aptitude. Teachers can take these abilities into consideration during instruction to maximize learning.

[18] Dimensions of Learning is an instructional framework developed by MCREL that identifies key elements of teaching and learning.

period, intermediate-level elementary teachers began voicing some concerns that the focus on Balanced Literacy was reducing needed support for comprehension and integrating literacy instruction with other topics, such as science and social studies.

High, Clear and Fair Academic and Behavior Standards

The decisions to be made here concerned how to establish and maintain standards for academic performance and behavior. Some schools interpreted this task to be one of adopting district policies and benchmarks for their own school; others interpreted it to mean developing mechanisms to create clear and consistent standards that students can understand.

Elementary Schools

There was little variation in this feature. All but two schools refer to the district's academic standards and benchmarks in their plans; all but four refer to the district's code of conduct; and all 14 elementary schools embrace the cooperative discipline approach.[19] Two schools describe the use of formative assessments[20] to set daily goals and instruction. And two schools mention the "building behavior rubric" and/or the use of a recovery room (e.g., time-out room) for students with behavior problems. (It should be noted that these two schools have the same SIF.)

Secondary Schools

All six schools mention adherence to the district's academic standards and benchmarks, but only four of the six schools mention use of the district's code of conduct. Procedures are specified for setting individualized standards within SLCs and for using such methods as instructional themes and assessments in two schools (Wyandotte High School and a Washington middle school).

Implementation Influences

As the implementation process moved forward, central office staff noted, especially at the secondary level, that there were some fairly entrenched low expectations for academic performance. This resulted in an increased district emphasis on the standards and benchmarks, and on the teaching and learning framework. The development of the online Curriculum, Assessment and Instruction Library (CAI) produced much discussion, as it involved developing rubrics for lesson plans that reflect the standards and benchmarks. While this task was still ongoing at the completion of this study, it had the effect of bringing the more generalized curriculum standards and benchmarks into the realm of daily practice through the example of these model lesson plans.

HOW MUCH WAS ACHIEVED IN IMPROVING INSTRUCTION?

We have data on instruction in the district from three sources. Classroom observations in the Wyandotte and Washington clusters, conducted from the 1998-1999 school year through the 2001-2002 school year included information on instruction.

Each of the approximately 250 classrooms in Wyandotte and Washington observed each year were coded for the dominant classroom activities, including the teacher supervision/facilitation approach, the dominant learner/grouping structure (i.e., one learner, dyads, small groups, large groups or whole class), and the academic activities in which students were engaged (i.e., active learning, passive listening, recitation, group discussion, worksheet or workbook, reading, writing, or not academically involved).

Data were also collected through student surveys on these implementation outcomes. The survey included measures on the type of learning activities students experienced in reading and math classes – specifically on their opportunities to work in teams, work on projects, and have connections made among their subjects and between their school work and their lives outside school. Survey measures of academic standards included questions on being provided models of work, meeting standards, teachers having high expectations and knowing what it takes to succeed academically. Questions on behavioral standards included students treating each other and adults with respect, adults treating students fairly, and rules being clear.

[19] The cooperative discipline approach to conduct standards is a philosophy of discipline relying on everyone in the community having the same standards, expectations and known consequences.

[20] Formative assessment is an ungraded assessment used to provide students and teachers feedback about each student's understanding of the curriculum to date.

Chart VII-1

Increase in Likelihood of Observing Small Grouping

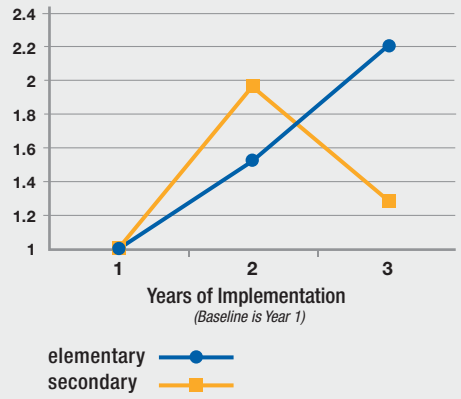


Chart VII-2

Decrease in Likelihood of Observing Large/Whole Group Instruction

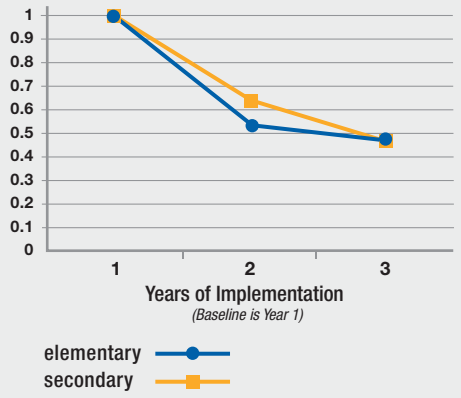
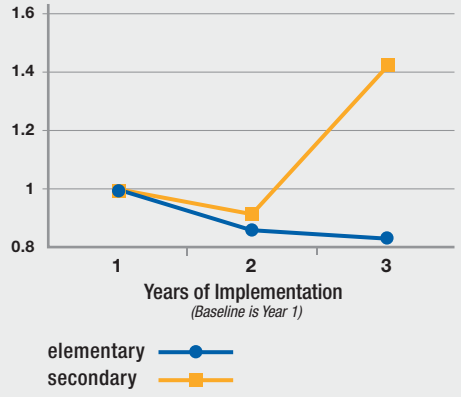


Chart VII-3

Increase in Likelihood of Observing Active Instruction



Small Grouping Strategies

In elementary schools, we were twice as likely to observe students being instructed in classrooms using small grouping strategies after three years of reform than after only one year of reform in the Wyandotte and Washington clusters (see Chart VII-1). So, for every 100 students instructed in this format in Year 1 of implementation, an additional 122 students were experiencing learning in small groups after three years of implementation in these two clusters.

The improvements in the secondary schools were also statistically significant, but not as large as in the elementary schools. In middle and high schools, there was a 28 percent improvement in the likelihood of students being instructed in classrooms where small groups were being used.

At the same time small grouping was more likely to be observed, we found significantly less instruction occurring in the predominantly large and whole group formats that are the hallmark of teacher-centered, passive instructional strategies. In both the elementary and secondary schools, the likelihood of observing this type of instruction decreased by about half (52% and 53%, respectively) (see Chart VII-2).

The relatively large improvements in small group instruction, especially at the elementary level, suggest that schools did effectively increase their use of small group instruction during core instructional times. This may be due to the implementation of such literacy initiatives as Just Read, Read 180, Balanced Literacy (Fall 1999) and the Read Aloud (Fall 2000) programs. The professional development around all these initiatives required staff to learn how to apply these techniques in small groups. The design of Balanced Literacy utilized a leveling strategy, which especially required smaller groups for students who were reading at lower reading levels; thus the larger effect on the at-risk group.

Active Instruction

During the classroom observations, active instruction was defined as actively engaging students in a structured way (e.g., dyads or other intentional grouping, with equal participation) to include any of the following: writing, discussing, sharing and/or manipulating learning equipment or materials.

Based on classroom observations, there were no significant improvements in the likelihood of seeing active instruction occurring in elementary classrooms between the first and third year of implementation in the Wyandotte and Washington clusters. However, in secondary schools, the likelihood of active instruction occurring increased by nearly half (43%) during this period (see Chart VII-3). There was also a 24 percent reduction in the likelihood of observing secondary classrooms where passive instructional techniques were dominant (e.g., lecture, seatwork, etc.) (see Charts VII-4).

Students also reported on their experience of the features of active and connected learning. Based on their surveys, elementary students were not significantly more likely to experience active learning in buildings with three or more years of implementation (see Chart VII-5). High school students, however, were nearly two-thirds more likely (64%) to report working in teams on projects that connected subjects and life outside of school. In other words, for every 100 high school students experiencing this type of instruction before implementation of FTF, about 164 were experiencing it once buildings had implemented the reform for three years.

There were no significant differences in the likelihood that students at all levels responded to the survey in patterns indicating they were still primarily experiencing traditional, passive forms of instruction (see Chart VII-6).

So, overall, the observation data show improvements at the secondary level in the move from passive to active instructional strategies and the survey data show high school students were more likely to experience active learning opportunities.

The lack of consistent improvements during the first three years of implementation in the occurrence of active instructional strategies in Wyandotte and Washington schools may be explained, in part, by continued concerns with behavior issues expressed by the teachers we interviewed throughout the five-year span of this study. Active instruction requires a certain amount of "letting go" of students (e.g., encouraging students to interact with each other in the context of group projects). We speculate that teachers with inadequate behavior management strategies would be reluctant to use active learning

Chart VII-4

Decrease in Likelihood of Observing Passive Instruction

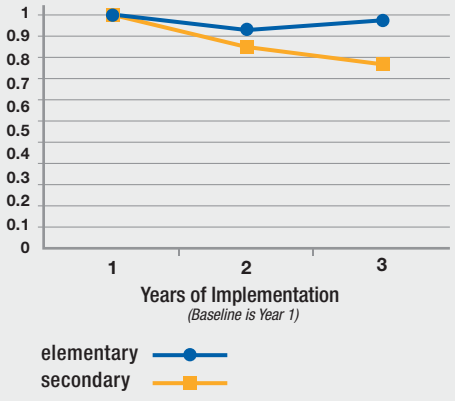


Chart VII-5

Increase in Likelihood of Active Instruction (Student Survey)

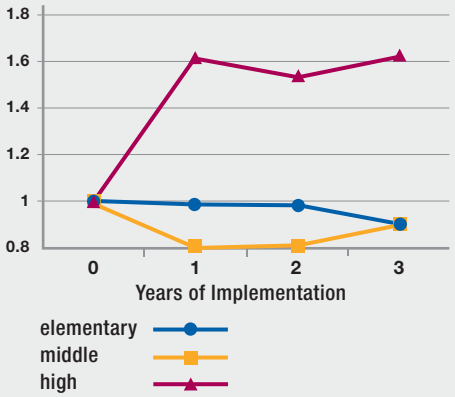
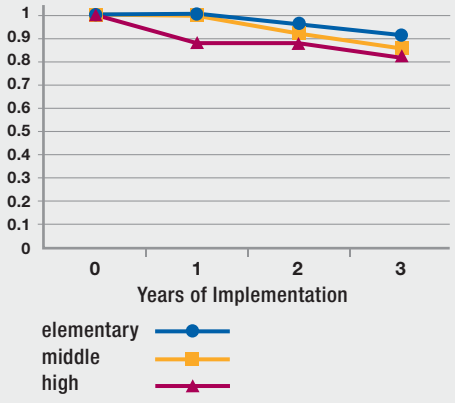


Chart VII-6

Decrease in Likelihood of Passive Instruction (Student Survey)



methods. Because the time span of this study included a period with teacher shortages and long-term substitutes, as well as an increasing number of less experienced teachers, the observed declines in active instruction might be reasonable. Furthermore, teachers in schools that experienced a greater enrollment of minority and low-income students due to the redrawing of school boundaries also expressed concerns about the behavior of "those" students. Their lack of experience in working with urban learners may have contributed to a greater reluctance to use active learning strategies.

It is also possible that the divergence in results shown by the different methods is meaningful. In order for classroom observations – which are conducted at a single point in time – to detect improvements in instruction, the desired techniques would need to be in practice on a regular basis. This would result in them being observed during whatever random class session was selected for the study. However, survey measures are framed in terms of when students are learning. Since students have the opportunity to reflect on the entire last year of instruction, these measures may be more likely to detect changes that have occurred over the entire school year, and not necessarily in any one class session. There is a convergence of both methods at the secondary level.

Academic and Behavioral Standards

Students at all levels were significantly more likely to experience high academic and behavioral standards in their schools after implementation of FTF. After three years of implementation in the Wyandotte and Washington clusters, 178 elementary students were experiencing high standards for every 100 that were having this experience prior to implementation (see Chart VII-7). Middle school students were over a third (39%) more likely to have high academic and behavioral standards in their schools after three years of implementation. The greatest gains were seen for high school students. Nearly four times the number of high school students (375 for every 100) in the first two implementing clusters were in the highest category of standards after three years of implementation than at baseline.

At the same time, more students were experiencing the highest levels of standards, and fewer students were experiencing the lowest levels of academic and behavioral standards. At both the elementary and secondary levels, decreases of between one-quarter and one-third of students per 100 in this lowest category occurred after three years of FTF reforms (see Chart VII-8).

The improvements in standards began after only one year of implementation. The line graphs show a steady increase for the elementary and middle schools in the high category of standards. High schools showed a slight decline in the first year of implementation followed by a steep increase in

Chart VII-7

Increase in Likelihood of High Academic and Behavioral Standards

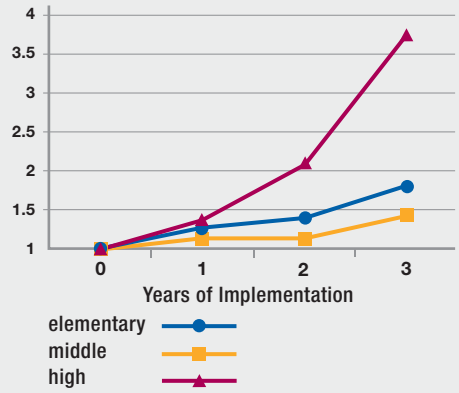
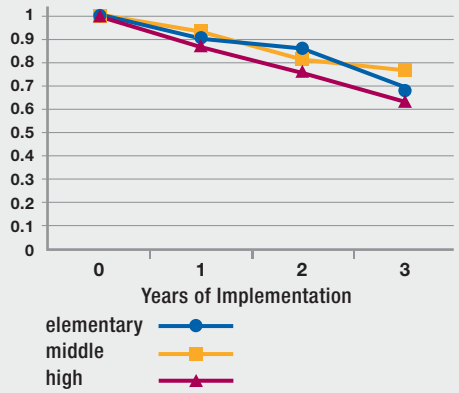


Chart VII-8

Decrease in Likelihood of Low Academic and Behavioral Standards



Years 2 and 3. The pattern of progress in moving students out of the lowest category of standards was steady across each year of implementation at all three levels.

The improved student perceptions of instructional and behavioral standards may coincide with improved clarity in those standards. During the period of this study, the district revised and published widely its standards and benchmarks, and began providing data to each school about student performance with respect to the standards, and developed a teaching and learning framework intended to implement those standards. It is reasonable that students would be able to perceive this increased clarity and be better able to perceive what those standards are. During the same time period, the district also encouraged schools to implement programs that provided strategies for improving student conduct and/or character. By 2000-2001, almost every school in the district was implementing some type of behavioral standards program (e.g., Positive Behavior Supports, Heart of a Champion), which may have made issues around treating each other in a fair, respectful manner more salient to students.

ELEMENTARY SCHOOL EXAMPLE

Instructional Implementation

High, Clear and Fair Expectations: A Washington cluster elementary school provides an example of the promotion of student perception of standards. With regard to behavior, there was a school-wide discipline plan in place that utilized a Positive Behavior Support model developed in collaboration with the University of Kansas. Posters focused on "character building" throughout the school. Academically, we observed high expectations, with professional development focused on helping teachers demonstrate teaching that reflected the district's standards and benchmarks. Hallway bulletin boards had benchmarks posted and examples of student work for that benchmark. Weekly grade-level meetings were devoted to the implementation of standards and benchmarks. Grade-level teams analyzed student performance data by grade and compared them to the standards, producing instructional plans for the coming week. The SLC teams, which also met weekly, coordinated these grade-level plans across the grades.

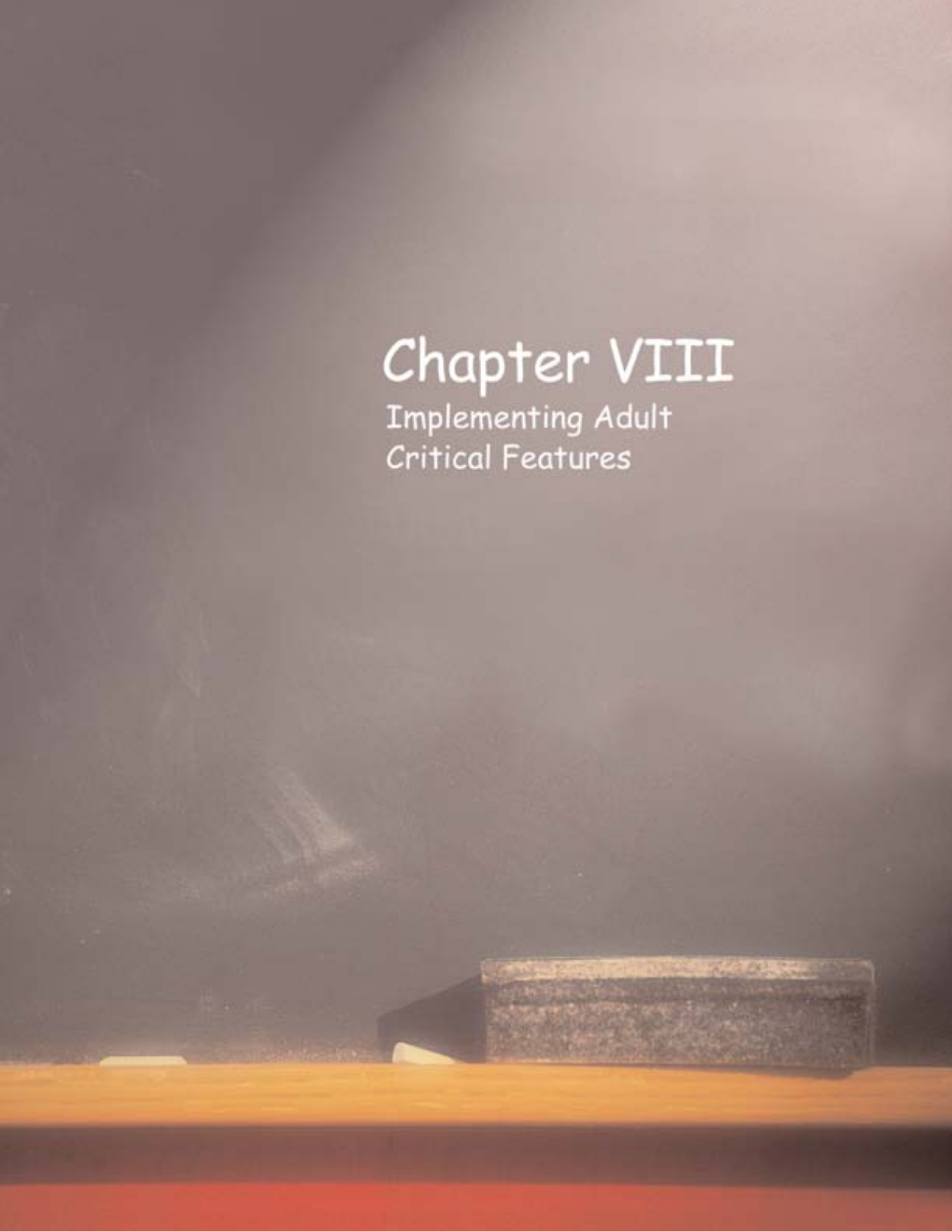
SECONDARY SCHOOL EXAMPLE

Instructional Implementation

Enriched and Diverse Learning Opportunities: The impact of teacher experience and confidence in behavior control on the implementation of active learning strategies is illustrated by two observations taken at a Wyandotte cluster middle school in the 2001-2002 school year. Each of the two teachers observed were engaged in the professional development offered to enhance their teaching strategies; in this particular school, the emphasis of the principal was on project-based learning. In one case, the long-term substitute makes no attempt at all to implement that strategy, while in the other case, the teacher is effectively engaging students in their projects. The very different outcomes in the classroom may be due to differences in experience and their expertise in classroom management:

Ms. O is a long-term substitute teacher for math class in Learning Community Two. She has been in this class since the beginning of the school year, but does not have her teaching certification. She is having a difficult time with students acting out, not following the rules and ignoring her requests. She watches carefully as students work to complete their worksheets. The students talk loudly about anything but math. One student, who received behavior checks earlier in the class period, walks around as he pleases visiting the other students. Several students chatter softly, occasionally lowering their tones to a mere whisper. Before the end of class, Ms. O reaches the end of her rope. She asks Mr. T, who is next door, for a referral sheet.

Ms. L's class in Learning Community Four busily works on their small group projects. They write their summaries on a white board in front of class. "This is the last report for student projects," she reminds them. Each group's science project has many parts, and each student in the group has to contribute to the final product. They must do as any good researcher would do, including defining the problem, developing the hypothesis, creating a design, identifying the variables, collecting data, analyzing data, reporting their findings and, finally, making suggestions for future research. As they check each other's work and compare their results with other groups, Ms. L strolls around the room encouraging her students to work together and provides enough guidance for them to be able to answer their own questions.



Chapter VIII

Implementing Adult Critical Features

MAJOR FINDINGS

- *Smaller elementary schools tended to perceive that they already had a sense of collective responsibility, since they routinely met informally and shared responsibility for students*
- *Collective responsibility appeared to be enhanced by the distribution of student performance data, and the development of such strategies as posting student work, word walls, and other mechanisms to enhance the school-wide standards and benchmarks..*
- *The shift in meaning of instructional autonomy and supports, from encouraging teacher autonomy to developing a broad repertoire of teacher skills, left some staff feeling confused about this feature in the middle years of implementation – this confusion abated in the last two years of the study.*
- *Principal leadership and the skills of SIFs made a difference in the quality of professional development and follow-through at different schools.*
- *Flexible allocation of resources was relatively easy to accomplish, compared with other features. All schools appeared to successfully delegate decision-making about space and supplies to the SLC. In addition, most schools maintained participation from the SLCs on school-wide, decision-making bodies related to hiring new personnel and budgetary matters.*

Due to modifications in the staff survey, we were unable to quantitatively track staff experiences of changes in the adult critical features of FTF after two years of implementation. However, we were able to track staff perceptions of whether the critical features overall were implemented (combining student and staff critical features), and this indicator did show significant improvement. But for the purposes of this section, we present only qualitative data for the Wyandotte and Washington clusters.

HOW WERE ADULT CRITICAL FEATURES IMPLEMENTED?

According to FTF, in order to meet the challenge of implementing and sustaining the student critical features, teachers need resources and supports for the new work they are expected to complete. The student critical features require that teachers work differently and relate to each other in new ways. Within Small Learning Communities (SLCs), adults must be able to decide what to teach, as well as allocate resources in response to their own and their students' learning needs. They must also take collective responsibility (with students and their families) to ensure that all students in their communities are successful. The staff critical features – collective responsibility, instructional autonomy and support, and flexible allocation of resources – focus on the adult responsibilities and supports necessary to ensure that all students' needs are being met.

Collective Responsibility

Here, the question was how to designate responsibility and accountability among all members of the staff for the performance of all students.

Elementary Schools

Almost all Wyandotte cluster schools interpreted this feature to mean that they were expected to state their target achievement goals. This may, however, be a result of the fact that the Wyandotte plans did double duty as the schools' Title I plans (which require statements of those specific goals for student improvement). In contrast, the Washington cluster plans suggest that these schools interpreted this feature to mean school-wide, grade-level coordination of assessments and curriculum, and/or goal setting and curriculum planning at the SLC level. This interpretation may be a reflection of the more explicit planning guidelines provided to Washington cluster in 1998. Four schools also added peer coaching or "buddy" systems for staff as a mechanism to achieve collective responsibility. Three schools included parents and/or students in this feature by specifying that parents and/or students would be given clearer ideas of their responsibilities. Finally, one smaller elementary school stated in its plan that it "already had" a sense of collective responsibility.

In practice, other small schools, while they did not specify this in their plans, also expressed a belief that they were already implementing collective responsibility in their schools. Indeed, the staff in smaller schools did appear to be more familiar with students in grades other than their own, and students did talk of knowing the other teachers in the school. Also, strategies intended to implement the "clear and fair" critical feature for students – such as posting on word walls student work and/or reading scores by grade levels in the hallways – seemed to foster a collective sense of both achievement and responsibility among the adults as well.

Secondary Schools

All but two of these six schools (two middle schools) designated SLC members as responsible for assessing and setting student goals. To facilitate this, mechanisms were specified for SLCs to have time for planning. Beyond that, a variety of unique strategies was mentioned (e.g., whole-school goal setting by the site council, reference to standards and

benchmarks, and collective instructional duties during power hour).

In practice, staff in some SLCs appeared to consider the role of the SLC coordinator to be similar to school administrators, in that students with disciplinary problems were sent to the SLC coordinator's classroom where she or he was expected to deal with the problem. For the most part, however, there was a sense of joint effort to maintain discipline, set goals and work together on behalf of students in the SLC.

Implementation Influences

The district began distributing student performance data by SLCs in 2000; this provided a concrete and public benchmark for the teams and may have increased a sense of collective responsibility. One of the observed barriers to enhancing collective responsibility was a tendency of staff to blame parents for students' poor performance. The implementation of the Family Advocacy program in 34 of the SLCs in schools throughout the district may be considered a strategy to enhance collective responsibility by improving partnerships between home and school.

Instructional Autonomy and Supports[21]

Not surprisingly, the initial school plans were focused on developing mechanisms to ensure that instructional decisions and decisions about support needs were in the hands of teachers. Later mechanisms were more oriented toward promoting an atmosphere of continuous professional development and maintaining a focus on instruction. As the implementation of FTF evolved, the critical feature of "instructional autonomy" was changed to "equip, empower and expect all staff to improve instruction." This revision to the critical feature reflected a move away from the belief that teachers needed more control over instruction and toward the idea that teachers needed to be provided the support necessary to improve instruction and be held accountable for doing so. A confluence of factors contributed to the reshaping of this critical feature, including the struggles the district was encountering in improving instruction, the enactment of the No Child Left Behind Act (NCLB), and an understanding that the district could more effectively use its resources to support instruction with a common curricula.

Elementary Schools

Ten schools specified that primary instructional decisions would be made in SLCs; four specified making decisions in grade-level teams. Coordination across SLCs for some school-wide planning was specified in three schools. All of the Washington and two of the Wyandotte cluster schools also interpreted this feature to mean that professional development was responsive to teacher-expressed needs, and/or mechanisms for determining those in-service needs would need to be laid out. Three schools specified that decisions would be based on individual student needs. Finally, one Wyandotte school noted that instructional autonomy meants "having a sense of empowerment and inclusion in planning"; there was no elaboration to explain how teacher empowerment would be operationalized.

Although the administration shifted its thinking away from "instructional autonomy" and toward "equip, empower and expect all staff to improve instruction," the exact name of this critical feature was not changed in this district because of an expressed need to maintain an appearance of consistency. As a result, some staff did not believe the central office was allowing them to implement "instructional autonomy," as they believed the feature was defined. Toward the end of the research, when staff began to perceive concrete results of changing practices in the classroom, this issue began to disappear across the 20 schools in the study. Particularly with the advent of new teachers who were attracted to the district because of the professional development opportunities, the image of the district's emphasis on development began to be a source of pride.

Secondary Schools

There was very little variation on this feature among the secondary schools. All stated that decisions about instruction would be made within SLCs; four described staff input in professional development plans (also to be developed within SLCs). One school described staff participation on school-wide committees (e.g., Budget) as part of instructional autonomy. (Note: There are other schools that establish school-wide committees, but they elected to describe these under flexible allocation of resources.)

In practice, secondary staff were similarly confused about the meaning of "instructional autonomy." As

[21] The name of this critical feature was changed to "Equip, empower and expect all staff to improve instruction" when FTF was expanded to other sites.

the emphasis shifted more toward professional development, the ability of the principal and/or SIF to serve as instructional leaders emerged more as a factor in the ability
of the school to implement this feature.

Implementation Influences
The district-level initiatives to provide comprehensive staff development and to focus on instruction (see section above) reduced the amount of variation in implementing this feature. All schools were involved in Wednesday early release in-service sessions; all schools had access to the literacy initiative and the instructional coaches; and all schools were impacted by the reorganization of executive directors to focus only on instructional improvement. In addition, principals were encouraged to (a) become instructional leaders, and (b) develop teacher leadership. Mechanisms for encouraging these shifts in leadership included a Leadership Academy offered at the University of Missouri-Kansas City and attended by some principals; monthly cluster meetings, including both principals and SIFs, where the focus was on instructional change; and a principal-peer mentoring system that involved pairing principals who supported each other in making instructional change and empowering teacher leadership.

Flexible Allocation of Resources

This feature involved specification of ways to make flexible decisions about the level and use of resources. At the central office level, this feature meant becoming more responsive to meeting the instructional needs of the schools as opposed to monitoring or developing standardized responses to the tasks of operating the district. For example, the Human Resources Department became more flexible over time in order to develop creative ways to meet schools' personnel needs. Similarly, budget allocations and decisions about support staff versus instructional staff became more a responsibility at the school level. At the school level, this feature was defined more as providing discretion and autonomy at the SLC level for decision-making about resources, space assignments and hiring decisions.

Elementary Schools
Six of the seven Wyandotte cluster schools specified in their plans that SLCs would have the discretion to plan use of space and schedules; none of the Washington cluster schools made this statement, despite the fact that all of these schools were divided into SLCs. Seven schools had some type of representation on school-wide budget, hiring, staff development or other committees.

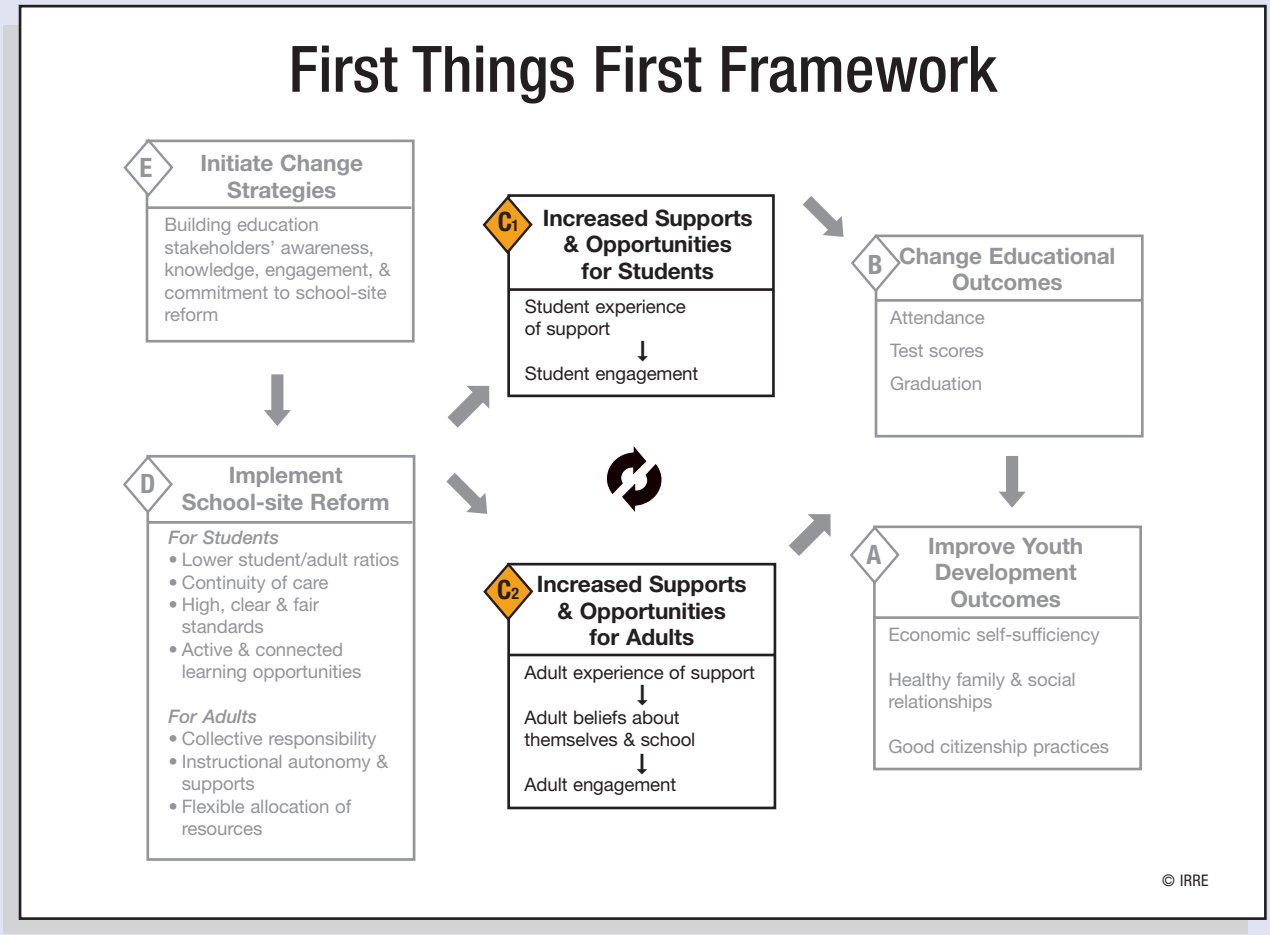
Secondary Schools
All six of the secondary schools stated that SLCs would have their own budgets for materials and staff development. Wyandotte High School and one Wyandotte middle school also stated that each SLC would have its own physical space and discretion over assignment of that space. All of the Washington cluster secondary schools and the other Wyandotte middle school specified that SLCs would make their own decisions about staff assignments. All of the Wyandotte cluster schools, but none of the Washington cluster secondary schools, stated that representatives from SLCs would sit on school-wide committees.

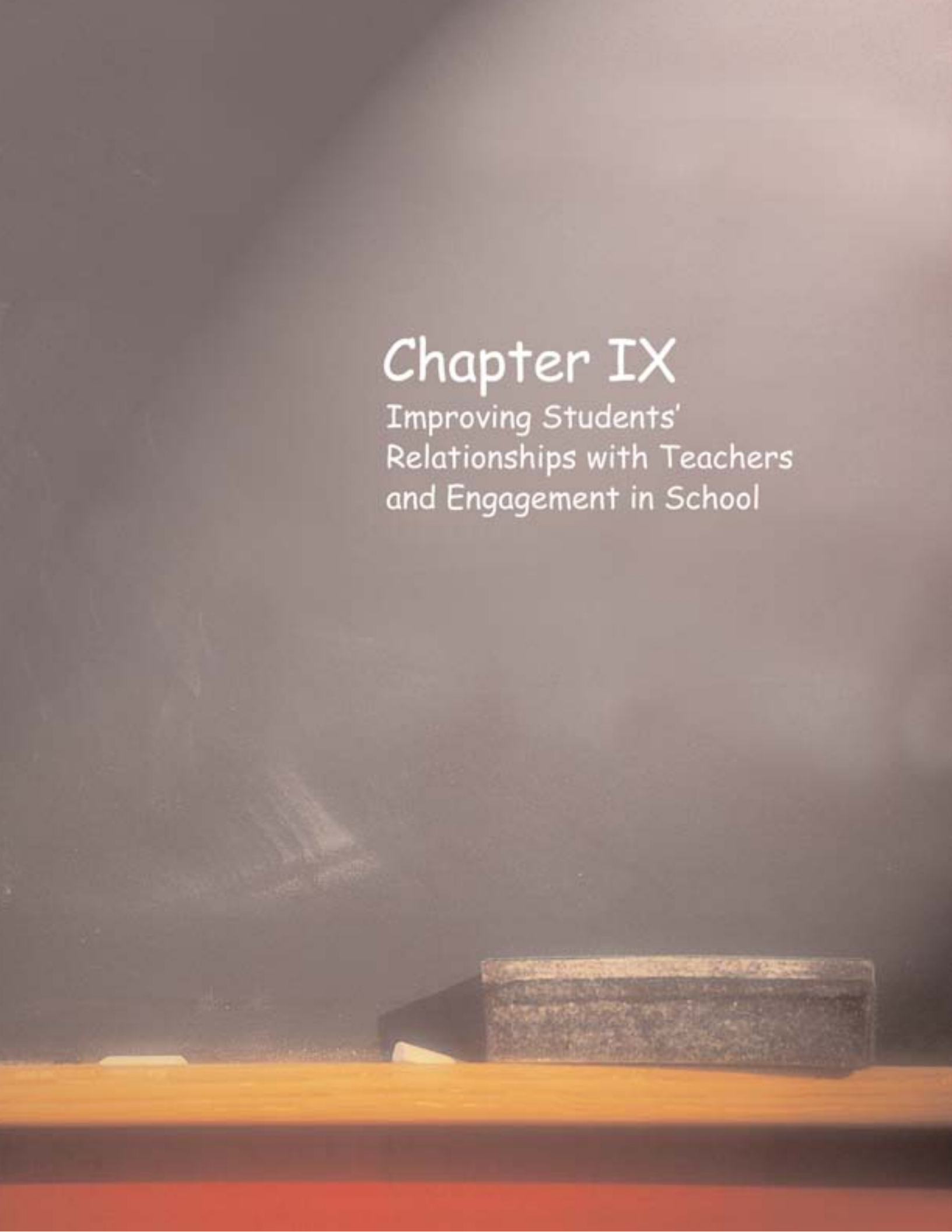
In practice, both elementary and secondary schools were able to implement this feature with relative ease in comparison to the other features. Principals received regular messages about the importance of "letting go" and developing teacher leadership. The most concrete example of "letting go" was the allocation of resources.

Implementation Variations
Physical layout placed restrictions on some schools' abilities to provide contiguous space for SLCs. SLCs typically had their own budgets for supplies. There were variations in the degrees of teacher involvement in school-wide committees, based on the leadership style of the principal. Involvement in the Principal's Leadership Academy appeared to facilitate greater levels of delegating power over resource allocation decisions, and using consensus-based decision-making.

INTERMEDIATE OUTCOMES

QUESTION 2: Did intermediate outcomes improve as buildings implemented First Things First?





Chapter IX

Improving Students' Relationships with Teachers and Engagement in School

MAJOR FINDINGS

- *Starting with the first year of implementation, there were significant increases each year in the number of students reporting they had relationships with their teachers that were characterized by support, caring, challenge and respect. These are the types of relationships that developmental research tells us lead to better outcomes for youth. The number of students in the lowest category of teacher support also decreased over the period of implementation.*

- *During implementation, significant improvements were made in the degree to which students were engaged in school. The greatest gains were made among students who started with the lowest levels of engagement (high school students), followed by middle school students. Elementary students were more likely than secondary students to be engaged in school prior to implementation of FTF; but even among this group, there were some gains in the proportion who were highly engaged in school after three years of implementation. There were also significant decreases in the number of students at all levels who were highly disaffected with school.*

Improvement in Student Intermediate Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Student Perception of Teacher Support	✔	✔	✔	✔
Student Engagement		✔	✔	✔

✔ ***indicates statistically significant improvement in outcomes***

Teacher Support:
Optimal = mean of 3.75 or higher on 8 items for elementary students; 3.25 or higher for secondary
Risk = Mean of 3.0 or lower on 8 items for elementary students; 2.5 or lower for secondary

Engagement:
Optimal = mean of 3.75 or higher on 10 items for elementary students; 8 items for secondary
Risk = mean of 3.25 or lower on 10 items for elementary; mean of 3.0 or lower on 8 items for secondary

Teacher support is one of the most important developmental supports associated with better outcomes for youth. Studies show that students who report caring and supportive interpersonal relationships in school:

- Have more positive academic attitudes and values, and are more satisfied with school (Solomon, Battistich, Watson, Schaps & Lewis, 2000; Felner et al., 1997; Shouse, 1996; Battistich, Solomon & Kim, 1995);
- Are more engaged in academic work (Ryan & Patrick, 2001; Marks, 2000; Solomon, Battistich, Watson, Schaps & Lewis,, 2000; Voelkl, 1995; Skinner & Belmont, 1993; Connell & Wellborn, 1991); and
- Are more likely to graduate and less likely to drop out (Croninger & Lee, 2001).

In turn, engagement has been linked with improved performance in school. Students who are engaged in school are more likely to earn higher grades (Willingham, Pollack & Lewis, 2002; Goodenow, 1993) and test scores (Willingham, Pollack & Lewis, 2002; Roderick & Engel, 2001; Lee & Smith, 1993, 1995), and have lower dropout rates (Croninger & Lee, 2001; Connell, Halpern-Felsher, Clifford, Crichlow & Usinger, 1995). In contrast, students with low levels of engagement are at risk for a wide range of long-term adverse consequences, including disruptive behavior in class, absenteeism and dropping out of school (Steinberg, Brown & Dornbusch, 1996; Lee, Smith & Croninger, 1995; Finn, 1993).

HOW MUCH WAS ACHIEVED IN IMPROVING STUDENT-TEACHER RELATIONSHIPS AND STUDENT ENGAGEMENT?

Both intermediate outcomes were measured through the student surveys administered each spring to all students in the third through twelfth grades. The measure of **teacher support** includes questions on how much students feel teachers care about how well they do in school, whether teachers have time for them, are fair, listen to them, and explain why they are learning certain things in school. The scale on **school engagement** includes questions on how hard students work in school, how important school is, if they come to class prepared and how they cope with challenge.

Teacher Support

After three years of FTF implementation, elementary students were 85 percent more likely to report high-quality relationships with their teachers than before the reform started (see Chart IX-1). There was a 60 percent improvement in the likelihood of having good relationships with teachers for middle school students, and a 82 percent improvement for high school students. So, for every 100 students who had good relationships with teachers prior to implementation, an additional 85 elementary students, 60 middle school students and 82 high school students received that level of support from their teachers after three years of implementation.

There was a corresponding drop in the number of students with poor-quality relationships (low support) with implementation. After three years of reform, the likelihood of feeling a lack of support from teachers dropped by one-third to one-half for students (elementary 42%, middle 33%, high 47%). (See Chart IX-2.)

Significant improvements in teacher-student relationships were seen from the first year of implementation. Across all years of the project, strategies for improving staff-student relationships were among the more well-accepted aspects of FTF, according to our qualitative interviews. This included the creation of SLCs (especially in the secondary schools) and looping (especially in the elementary schools). As early as 1998-1999, respondents were describing perceived improvements in relationships

between staff and students, and these reports remained constant throughout the study. Consistent with these comments, we observed teachers in team meetings discussing students and their particular home situations as a part of making decisions about them, interactions with students demonstrating caring attitudes, and extra help provided to students facing challenging life situations.

Engagement

After three years of implementation in all four clusters, the likelihood of elementary students being highly engaged in school had not change significantly (see Chart IX-3). However, secondary students (middle and high school) showed great increases in engagement after the reforms were in place for three years. The differential in relative improvement for this outcome results from the stark differences in the starting points at the different levels (see Charts IX-5 and IX-6). Prior to the FTF initiative, one-third of elementary students were highly engaged in school, whereas only 1 percent of middle and high school students fell into this category. The noteworthy trend here is that all of the improvement in the likelihood of students being highly engaged occurred after three years of implementation.

In contrast, the movement of students out of the lowest category of engagement showed steady improvement with each additional year of implementation for elementary and secondary schools. After three years of implementation, elementary students were 28 percent less likely to be disengaged from school, middle school students were 28 percent less likely and high school students were 29 percent less likely. This means that for every 100 disengaged students prior to FTF, after three years of implementation there were only 71 or 72 students in this category (see Chart IX-4). This result is not uncommon in education research. Studies have shown that outcomes sometimes show movement of students out of the lowest levels before increases are seen in the proportion of students in the top level (Snipes, Doolittle & Herlihy, 2002).

Chart IX-1

Increase in Likelihood of High Teacher Support

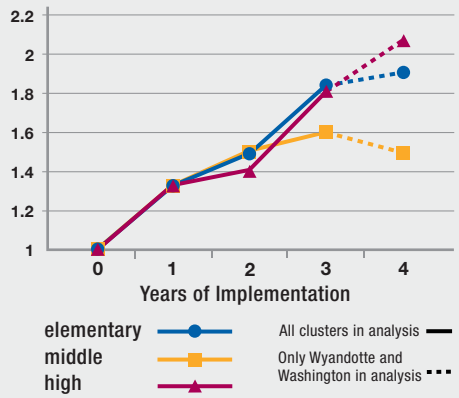


Chart IX-2

Decrease in Likelihood of Low Teacher Support

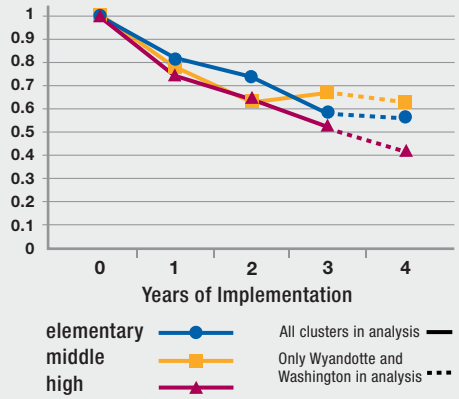


Chart IX-3

Increase in Likelihood of High Student Engagement

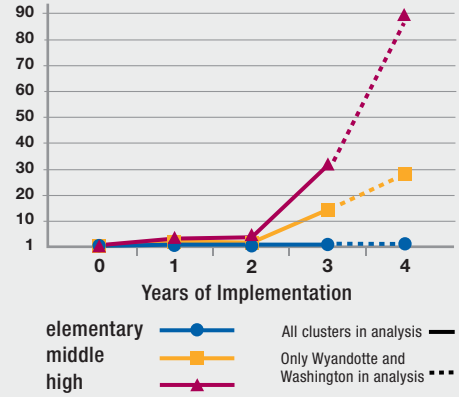


Chart IX-4

Decrease in Likelihood of Low Student Engagement

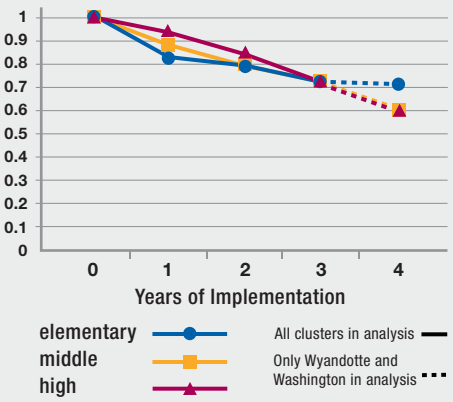


Chart IX-5

Mean Percent of Students With Optimal Engagement

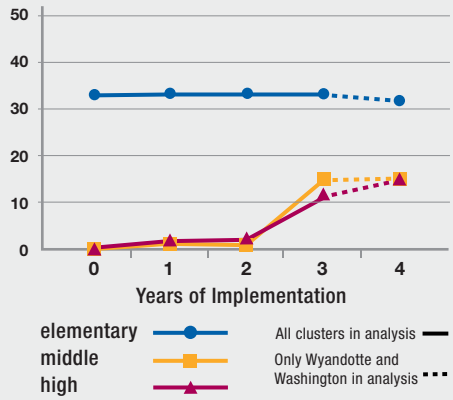
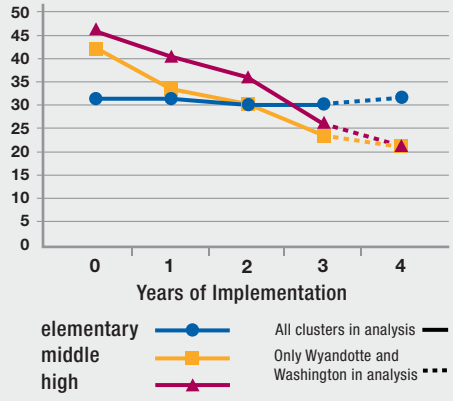


Chart IX-6

Mean Percent of Students with Low Engagement



The District and the Institute for Research and Reform in Education (IRRE) leadership were aware that engagement was showing slower change in the high category than was teacher support. Starting in 2000-2001, professional development for principals, SIFs and executive directors included a focus on student engagement during "walk-throughs." Training for walk-throughs included reaching consensus around what an engaged classroom would look like (e.g., students paying attention, asking questions, focused on the work product, etc.). These walk-throughs provided more accountability because they helped district administrators to better understand how, and whether, classrooms were engaging environments. At the same time, they signaled to teachers that engagement was a district priority. SIFs (now instructional coaches) also received additional training on engaging teachers in dialogues about active engagement; and techniques for assessing and enhancing student engagement during class time. This was a mechanism to support those teachers who needed additional help implementing the active engagement strategies they were learning in professional development and early release sessions.

ELEMENTARY SCHOOL EXAMPLE

Engagement

The internal process of instructional coaching and the focus on engagement is illustrated in this section by an interview with a SIF, conducted in Spring 2001. At that time, SIFs had just been informed that the next fall they would work exclusively with one school rather than two or three as in previous years. In this section of the interview, she is discussing her role in working with staff at an elementary school:

"Next year…I will have [a more] intense focus on the culture of this school. I [already] have an understanding of the culture of it. Currently, I am not here through all of the business of the day, so that is going to be different. I spend a lot of time here already, though. Teachers sometimes see that as an advantage, and sometimes it is like 'oh no, here she is – all day long an opportunity to come into my classroom to check my performance'…I am learning how to make an entrance into the classroom without being intrusive, but yet knowing there are certain things I am looking for…"

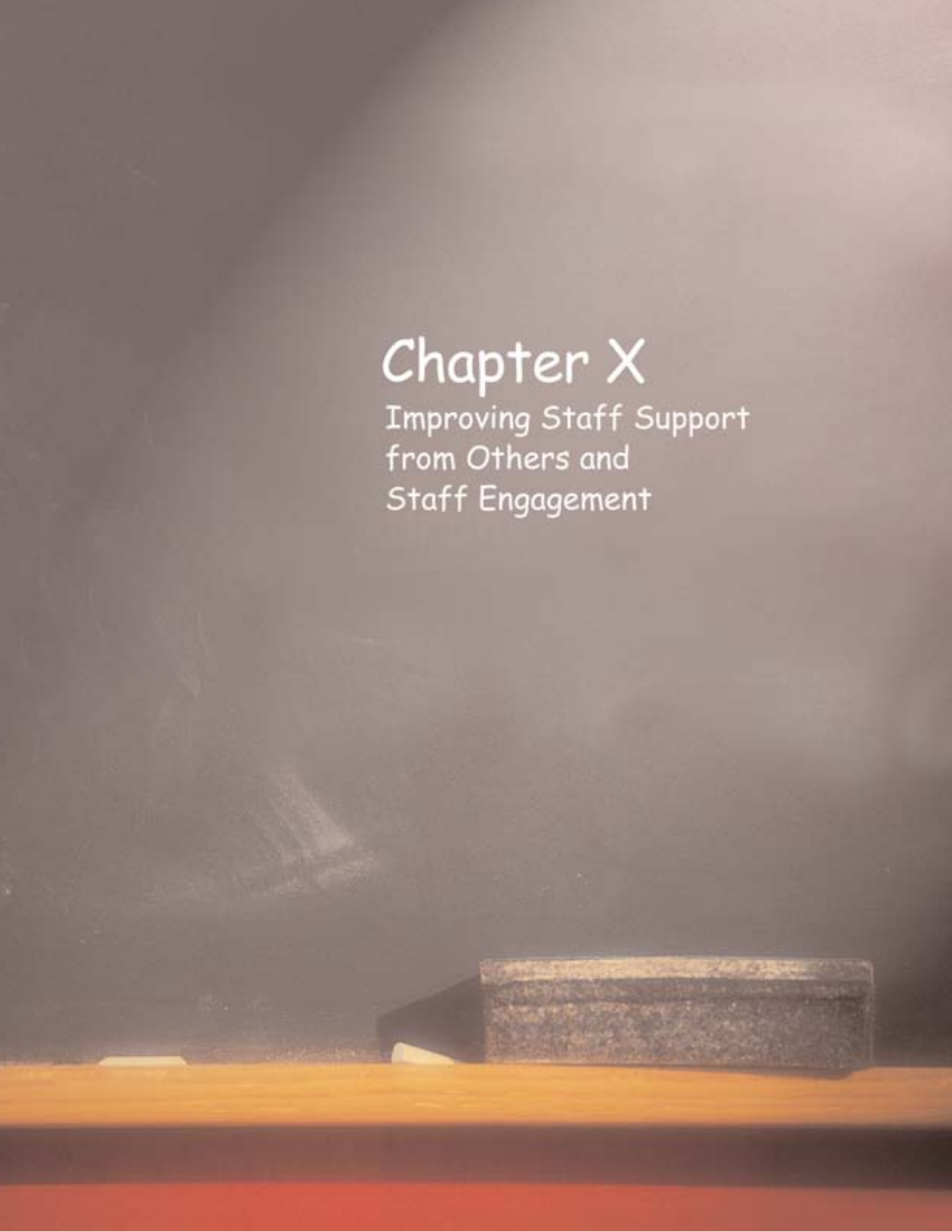
"There's a high level of awareness about student engagement. Some are aware of the mechanical part of it. There is an awareness to the level that when I am there they are trying to think of a structure to pull it off. [The next step after engaging the students] is looking at the outcome of what engagement is about. We have witnessed a strong engagement structure, but the quality of the work is still not really there. We are looking at student work. What is it saying about my performance as a teacher? We are also pushing for reflection and action. When I am teaching the lesson, what has happened in the moment that I can think on quickly to move to the right application. I am seeing that help more."

SECONDARY SCHOOL EXAMPLE

Teacher Support

A prevailing norm at one high school, propagated through the SIF, is that the primary purpose of the SLC is to foster high-quality relationships with students. Student interviews suggest that even those students who are disgruntled still strongly identify with their community and experience a sense of support from their teachers. Observations of teacher and student interactions in the health community suggest that the teachers and students enjoy an easy, relaxed relationship. The norm of caring as the key to success is reflected in a comment from the health community coordinator: "Test scores have gone up. I don't know, we would like to say good job but the difference is caring about kids…kids are steadily moving up. I have seen more kids trying. This is about pride in your community."

The health community coordinator is Hispanic and is able to identify with her students. She related how she had been alienated as a student, and what helped turn her around was her high school teacher. There is a strong belief in the value of relationships as a vehicle for helping students become engaged with their school. We observed the coordinator circulating in the room, touching and encouraging students, engaging students who were sleeping, etc. The other teachers we observed also appeared to have positive relationship with the students. The atmosphere was relaxed, and students appeared to feel free to approach and discuss things with the teachers. One teacher, for example, was good natured about teasing from students who told us they thought she had gotten much better since the beginning of the year and thought it was because she had benefited from the Wednesday afternoon professional development. The coordinator described advocating with the school administration on behalf of a student who had made great progress the previous year. This year, however, the student's mother had remarried and the student was experiencing stress at home with the readjustment of her relationships, which was showing up at school in the form of behavior problems and tardiness. The coordinator did not want this student to be suspended because she thought being away from school would exacerbate her problems . "It'll just destroy everything we were working toward." The SIF noted that prior to implementation of FTF, most teachers would not have known these details about a student's life outside school.



Chapter X

Improving Staff Support from Others and Staff Engagement

MAJOR FINDINGS

- *After three years of implementation in the Wyandotte and Washington clusters, the only area where significant improvements were seen in support from colleagues was in the proportion of secondary staff who fell into the lowest level of this outcome.*

- *However, there were improvements across the board in staff's perceptions of the degree of support they received from building and district leadership (system support). After FTF had been implemented in buildings for three years in the Wyandotte and Washington clusters, staff at both the elementary and secondary levels were more likely to perceive support from system leaders at the highest levels, and significantly fewer staff fell into the lowest level of perceived support from system leaders.*

- *The degree to which staff were engaged in their work increased in two ways after three years of FTF implementation in the Wyandotte and Washington clusters. The likelihood of staff feeling highly engaged in their work increased significantly; and there was a much lower likelihood of staff being disengaged from their work.*

Improvement in Staff Intermediate Outcomes in Wyandotte and Washington Clusters

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Colleague Support				✔
System Support	✔	✔	✔	✔
Staff Engagement	✔	✔	✔	✔

✔ ***indicates statistically significant improvement in outcomes***

Colleague Support:
Optimal = mean of 3.5 or higher on 6 items
Risk = mean of 3.0 or lower on 6 items

System Support:
Optimal = mean of 3.5 or higher on 10 items
Risk = mean of 3.0 or lower on 10 items

Staff Engagement:
Optimal = mean of 3.5 or higher on 12 items
Risk = mean of 3.0 or lower on 12 items

The FTF reform model posits that school staff need support from their colleagues and administrators to enable them to meet the challenges associated with implementing the student critical features. Studies of teachers and their work show that factors promoting their engagement are similar to the ones that work for students. Specifically, the quality of teachers' relationships with their students and each other, along with the diversity, flexibility, challenge and recognition associated with their work, foster engagement, willingness to innovate and productivity (Wohlstetter et al., 1997; Lee & Smith, 1994, 1996; Little, 1982; Little & McLaughlin, 1993; Louis, 1992, 1995; McLaughlin, 1993; Rosenholtz, 1989).

HOW MUCH WAS ACHIEVED IN IMPROVING STAFF RELATIONSHIPS AND ENGAGEMENT?

Staff's perceptions of the amount of support they received in their work and the degree to which they felt engaged in their work was measured through staff surveys. The surveys were administered each spring to all staff in each building. However, the staff survey was discontinued after the 2001-2002 school year. The survey form was significantly shortened in the 2000-2001 school year, but district leaders continued to hear complaints from staff about completing surveys each year (the staff surveys began in 1997-1998). As a result, staff intermediate outcomes were only measured through three years of implementation in the Wyandotte and Washington clusters (2002-2003 was Year 3 in the Harmon and Schlagle clusters). Therefore, results reported here are only for the first two implementing clusters.

Colleague support was measured through questions about whether staff encourage each other, share resources and go out of their way to help each other. **System support**, or support from leaders, included questions in two areas: whether staff believed building administrators supported them in doing what they need to do, in making their own decisions about students and in getting what they need from the central office; and whether the central office gave them the resources they need to do their work, supported educational innovation and provided professional development.

Staff engagement was measured with 12 questions on individual engagement and three questions on collective engagement. Individual engagement includes questions on how staff feel when teaching, if they feel they are just putting in time, and how they handle challenges. Collective engagement questions include perceptions about whether staff in the building, in general, do what is necessary to get the job done right, do not give up in the face of difficulties and go beyond the call of duty.

Colleague Support

The degree of support staff felt they received from their colleagues did not change substantially over the course of implementation in the Wyandotte and Washington clusters (see Charts X-1 and X-2). The one area where a small, but significant, improvement occurred was in the likelihood of secondary staff feeling low levels of support from their peers. After three years of implementation, this group was 16 percent less likely to perceive their colleagues in this way (Chart X-2). The trend lines show that it took two years of implementation for secondary levels of low support to decline after increasing in the first year of implementation. The likelihood of elementary staff believing they received low levels of support from colleagues increased in Year 1 then started to decline, but had not significantly improved after three years of implementation.

System Support

Elementary and secondary staff were about 40 percent more likely to feel supported by building and district leadership after three years of implementation (42% and 45%, respectively) (see Chart X-3). There were also small, but significant, decreases in the likelihood that staff felt unsupported by leaders after three years of reform (Chart X-4).

In both the elementary and secondary levels, the likelihood of staff feeling highly supported by leaders declined before it began to improve – for the first two years of implementation, for elementary staff and for the first year in secondary schools. The improvements in staff feeling low levels of support started after two years of reform for both levels.

Staff Engagement

For every 100 elementary staff who were highly engaged in their work prior to implementation in the Wyandotte and Washington clusters, 136 felt this same way after three years of implementation (see Chart X-5). Secondary staff showed a 51 percent improvement in the likelihood of feeling highly engaged. There were also improvements in staff feeling disengaged from their work. Elementary staff were 14 percent less likely to be disengaged and secondary staff were 26 percent less likely to be disengaged than they were prior to FTF implementation (Chart X-6).

Unlike the support variables, the improvements in engagement began after only one year of implementation. However, for secondary-level staff, the likelihood of feeling engaged rose in both Year 1 and Year 2, but dropped in Year 3 – although it remained significantly higher than at baseline.

It is not unusual in reform efforts to see what is commonly referred to as an "implementation dip," where outcomes get worse before they begin to improve. This was the case with the support outcomes for staff, but not for staff engagement.

Based on qualitative data, the improvement in staff perceptions of support from leaders and engagement in their work was related to the quality of principal and SIF leadership. It appeared that staff enthusiasm and acceptance for the reform was bolstered in schools where the principal was hands-on, instruction-oriented and willing to build consensus about plans for changes and involvement in professional development. Also, the skills of the SIF, in terms of modeling collaborative behavior and facilitating professional development, seem to have contributed to the staff's engagement and support. (These links will be explored further in Chapter XII on relationships between building factors and improvements in outcomes.)

Chart X-1

Increase in Likelihood of High Colleague Support in Wyandotte and Washington Clusters

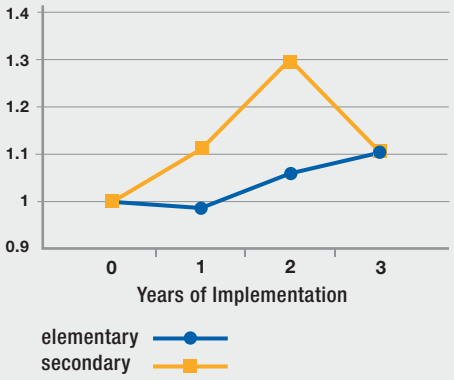


Chart X-2

Decrease in Likelihood of Low Colleague Support in Wyandotte and Washington Clusters

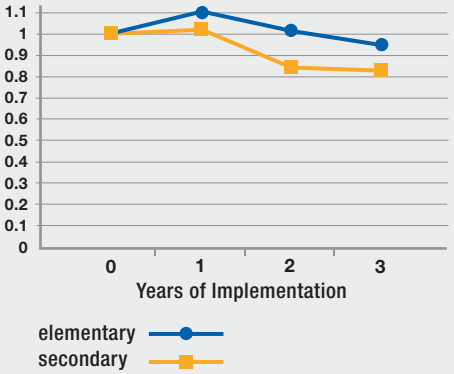


Chart X-3

Increase in Likelihood of Optimal System Support in Wyandotte and Washington Clusters

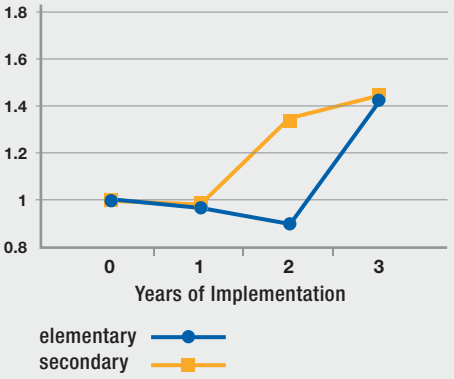


Chart X-4

Decrease in Likelihood of Low System Support in Wyandotte and Washington Clusters

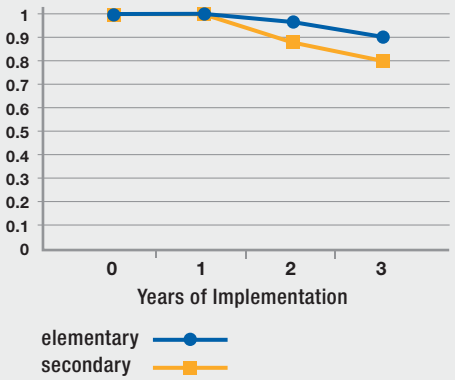


Chart X-5

Increase in Likelihood of Optimal Staff Engagement in Wyandotte and Washington Clusters

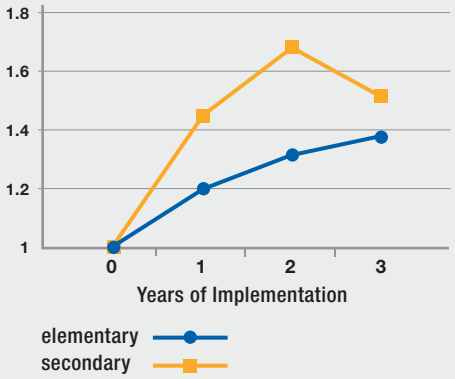
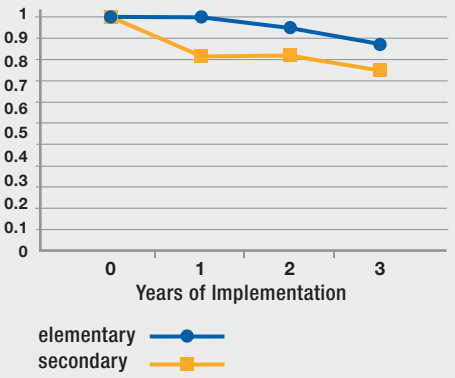


Chart X-6

Decrease in Likelihood of Low Staff Engagement in Wyandotte and Washington Clusters



However, strong leadership may be a necessary, but not sufficient, factor for improving support for staff and their levels of engagement. Support for innovations may be a relatively stable trait in school climate and organizational culture, which may be difficult to affect. Or, it may be affected by changes in key teacher leaders. We observed changes in the overall school climate in cases where a teacher (usually a long-time veteran) who was resistant to FTF resigned or retired. Due to the visibility of these veteran teachers as leaders within their schools, their retirement may have influenced teachers' perceptions of the degree to which their colleagues, overall, supported the reform.

The small, but consistent, improvements in perceived system support illustrate shifts in perceptions of the intent and value of supports offered by the central office. In the early years of the reform, mandates for structural change and specific training were perceived as interference and control. As time went on, however, these supports were increasingly viewed more positively. This may have been due, on the one hand, to the increasing perception that FTF was having an impact and, on the other hand, to the gradual resignation or retirement of those who resisted the reform.

Specific central office actions, which may have affected perceptions of system support, included the Wednesday early release times (which gradually became more popular), increased visibility of central office staff in the buildings (including the superintendent and executive directors, as well as the SIFs), and evolving reorganization efforts in the central office (which staff saw as evidence that the central office was willing to change as much as it expected the buildings to change). The approval of a bond issue to provide air conditioning and technology in all buildings district-wide (which occurred in Spring 2001) appeared to be a tangible morale booster.

In addition to these direct actions, steps taken to resolve problems at the district level may have removed sources of irritation and enhanced the perception of system support for the reform. These included (a) more flexible allocation of funding to enable the schools to have more autonomy and more equitable access to resources; (b) improved recruitment and streamlined hiring procedures in the Human Resources Department; and (c) improved software and other systems, as well as (relatively) improved troubleshooting responsiveness of the Technology Support Department.

The increase in staff engagement in the reform may be attributable to increasingly common perceptions that FTF was having an impact. In the early years of this study, respondents mostly described impacts in terms of the impacts on their school structure and relationships with colleagues and students. In the last years of the study, we heard even more comments about improvements in student performance. Several respondents in our final interviews referred to the "snowball effect" of positive results in student test scores, which they believed stimulated increased commitment to embrace the structural and instructional elements of the reform.

Similarly, the positive media attention about FTF that occurred in the last two years of the study may have affected staff engagement. Respondents commented about feeling a sense of pride in their district and in being part of this "exciting" initiative.

ELEMENTARY SCHOOL EXAMPLE

Staff Support and Engagement

An elementary school in the Washington cluster provides an example of a school where staff support and engagement was strong from the beginning and continued throughout the project, and demonstrates how the reform was able to capitalize on these strengths. This school had a strong principal who retired midway through the study, but was replaced by another, equally strong, principal for the last two years of the study. The SIF was also highly respected as a competent facilitator, and the staff were viewed as stable, competent and (for the most part) harmonious. The acceptance of FTF that we observed over the course of the study may be due, in part, to the very legitimate and authentic role that the Stakeholder Committee played from the very beginning of the planning year, 1998-1999, through the implementation of SLCs and facilitators. This acceptance was sustained during the transition from one principal to the next, regardless of the fact that these two principals have very different leadership styles. Also, the positive teachers at the school, along with the SIF and literacy leaders, remained hopeful and stayed focused on doing what was in the best interest of the students.

One challenge for the new principal was to change the stories told in the school about "those" kids. The principal used some Wednesdays to help staff understand what the expectations are for the entire building and then giving them support. Wednesdays, for the most part, were focused on active engagement and reflection of teaching practice. With the departure of the SIF at the beginning of the 2001-2002 year, the principal reported that she did not have the support and camaraderie she was used to with the SIF, but said that she stayed the course and kept the focus on teaching and learning.

Overall, there have been high norms for professional development, which began with the professional development work group, the SIF and the principal. Input was gathered from all staff members and many teachers reported attending workshops and conferences during the course of the four-year study at this school. During the 2001-2002 year, the principal and new SIF planned with SLC leaders to provide instructional support for SLCs. The four new teachers were involved in the New Beginners Teachers Program during 2001-2002. A decision was made to have one pilot teacher from each grade level provide training for math investigations and a KALL teacher was added to the staff. All of these positions continue to support the growth of the teaching staff at this school.

During the 1999-2000 school year, when the staff had to study the school's data, there was much blaming and wringing of hands about the low test scores, but then, according to the principal and SIF, the staff decided to focus on Balanced Literacy and developed a strong buy-in for this literacy program. At the end of 1999-2000, four teachers retired. Two of the retirees were described by respondents as positive about FTF and two were described as "blockers." As a result of FTF, some respondents said that teachers, overall, were communicating more and the negative staff members were being isolated. The working relationship between the old principal, the staff and the SIF was described as very positive. The staff started feeling very positive about opportunities to collaborate and work together. During the planning year, the high trust level among staff members was cited by respondents as being a positive force in countering the stress of changing demographics and occasional negativity on the part of some staff members. While everyone has not always been happy about certain decisions, there is a belief that everyone is listened to, options are explored, and though the majority agree with the decisions, everyone agrees to support them.

In summary, staff support and engagement at this elementary school may be attributed to (a) a strong principal and SIF leadership, (b) a positive school climate focused on instruction, (c) decision-making processes based on consensus-building and trust, and (d) the eventual departure of those "blockers" who were resistant to the initiative.

SECONDARY SCHOOL EXAMPLE

Staff Support and Engagement

One high school provides an example of the dynamics of teacher disengagement and the time and effort required to begin making changes in an entrenched school culture. Based on data from Fall 2000, the 74-person faculty included 21 teachers, or 28 percent, who had tenure of 25 or more years at the school. In contrast, the leadership team (both principals and assistant principals) turned over three times during the course of this study. Not surprisingly, given this dynamic, the culture was characterized by both teachers and administrators as a "them/us mentality." Mix in the tendency of some of these veteran staff to be vocal, negative and powerful, and the result was a high degree of resistance. The staff fought the concept of SLCs from the first year of planning by having lower and upper division SLCs (i.e., SLCs consisting of grades 9-10 and 11-12, respectively). Initially, the teachers decided to not have vertical teams (grades 9-12 in each SLC), because the respondents said they did not want to be like another high school in the district – they wanted to be unique and craft their own plan. Some respondents admitted that they did not want to teach freshmen and sophomores, so they supported the lower division and upper division plan. The SIF stated that it took a two-year cycle of students for the lower division teachers to see the benefits of looping and realize they wanted to keep their students four years.

The SLC coordinators in the lower division asked for assistance in developing leadership, team-building and facilitative skills. During the 2000-2001 year, through a New American School grant, they were offered a graduate-level course on these topics through the University of Missouri - Kansas City (UMKC). Their graduate credit was paid for out of the grant and the class was held at the school site during winter semester 2001. This course was a result of the work of the SIF. Half of the staff (all lower division teachers plus several electives teachers) chose to take the semester-long class, which was held at the school and focused on building stronger SLCs.

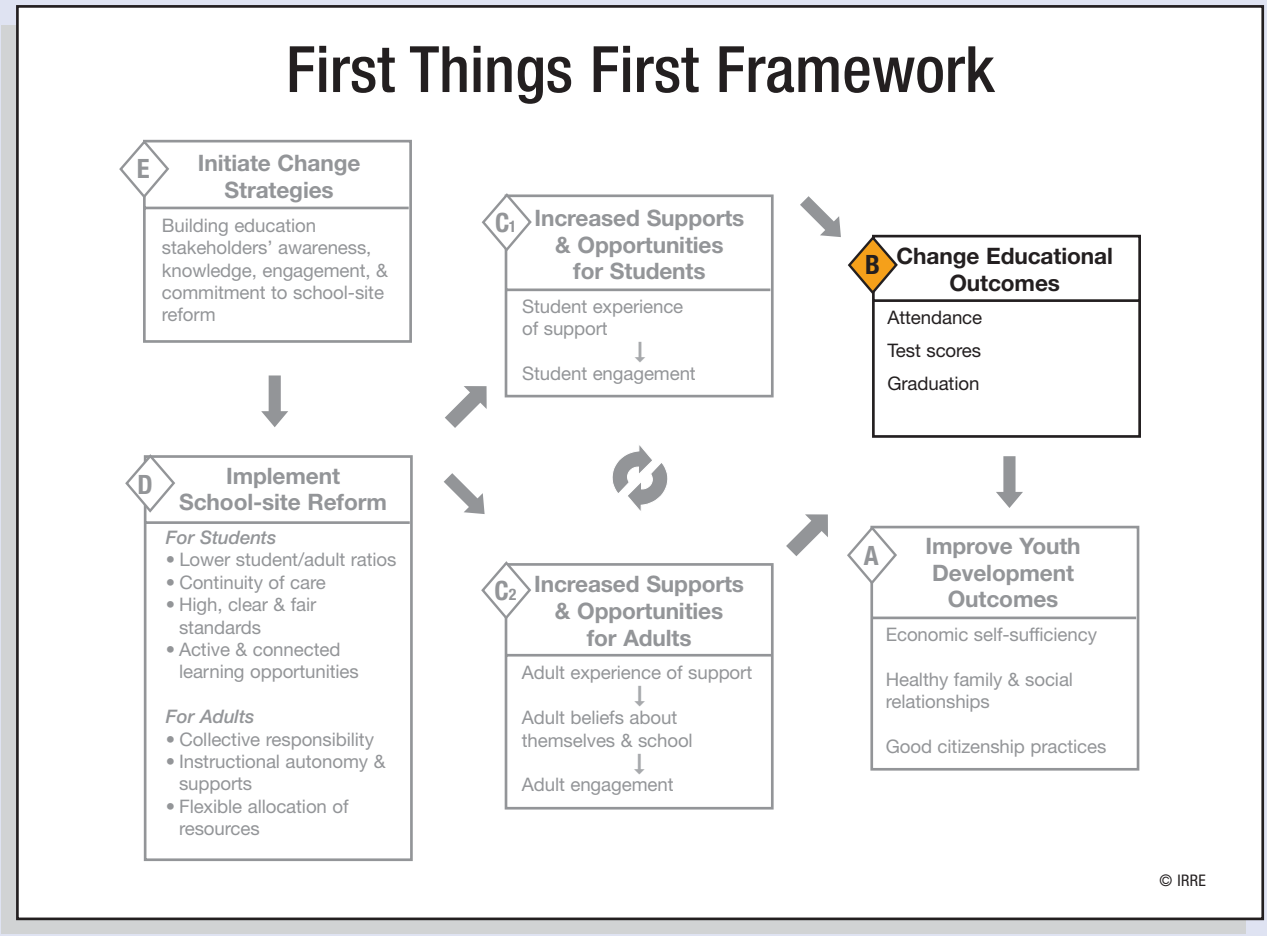
Those teachers and administrators who took the UMKC course evaluated the class as very helpful, and the lower division SLC coordinators said that they were using the skills they developed in their SLCs. In contrast, some teachers and administrators reported that some of the SLC coordinators in the upper divisions not only did not participate in the training but also did not push their SLC to reach its potential because they were fundamentally opposed to FTF.

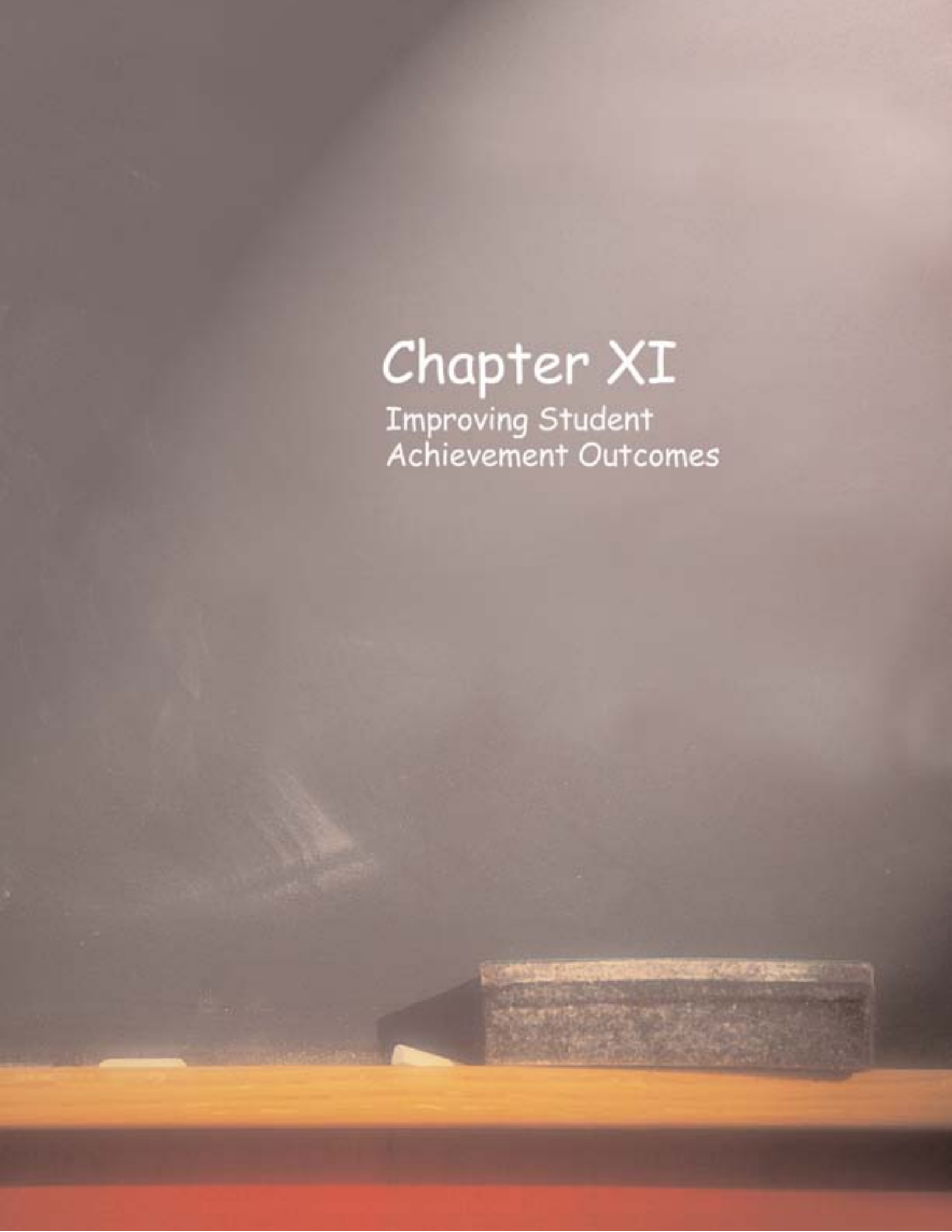
During the 2001-2002 year, there was a new administrative team and a mandate from the central office to begin planning to convert the SLC structure to vertical, theme-based units. The new administrative team that came on board at the beginning of the 2001-2002 year must be credited with enduring the "abuse" of this group of teachers and keeping a positive focus on the implementation of vertical SLCs. At the end of his first year, the principal said that it has been extremely difficult changing the culture from a them/us mentality, but his administrative team has not given up. He concluded the interview at the end of the 2001-2002 year by saying that 2002-2003 would be a critical year at his high school. In addition to implementing vertical SLCs, uniforms would become mandatory, based on a survey of students, staff and parents. Another big change was that eight teachers – many of whom were among the most resistant – retired or moved to other buildings at the end of 2002.

This example describes the combination of efforts required to "turn the ship around" in terms of staff engagement and colleague support: First, the lower division teachers saw tangible evidence of the benefits of looping. This led them, second, to ask for support and professional development on building quality SLCs. Third, mandates from the central office and a stronger administrative team clarified the expectations and direction of the school. And, last, those remaining teachers who had been resistant finally retired or moved to other assignments.

IMPROVING STUDENT ACHIEVEMENT OUTCOMES

QUESTION 3: Did outcomes for students improve as buildings implemented First Things First?





Chapter XI

Improving Student Achievement Outcomes

MAJOR FINDINGS

- *The likelihood of students meeting or exceeding district attendance standards increased significantly during the implementation of FTF; and the probability of students falling below district attendance standards decreased with each additional year of implementation..*
- *In general, the likelihood of students scoring proficient or above on both the state reading and math tests increased with additional years of implementation, while the likelihood of students scoring unsatisfactory on these tests declined.*
- *Students in the graduating class of 2002-2003 were significantly more likely to graduate than were students in the class of 2000-2001. That is, graduation rates would have increased from about 90 percent to more than 95 percent over this three year period.*
- *The likelihood of students dropping out of high school decreased by more than 90 percent in each of the four high schools during the course of FTF implementation. That is, schools with dropout rates in the range of 45 percent before FTF implementation would have declined to dropout rates around 2.5 percent after four years of implementation.*

Improvement in Achievement Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Attendance	✔	✔	✔	✔
State Reading Test	✔	✔	✔	✔
State Math Test	✔	✔	✔	✔
Graduation	NA	NA	✔	NA
High School Dropout Rate	NA	NA	NA	✔

✔ ***indicates statistically significant improvement in outcomes***
NA ***indicates not applicable***

Attendance:
Optimal = 1 or fewer absences per month
Risk = 1 or more absences per week

State Tests:
Optimal = Proficient or Above
Risk = Unsatisfactory

Graduation:
Optimal = Graduates within 5 years

HOW MUCH WAS ACHIEVED IN IMPROVING STUDENT OUTCOMES?

Student achievement and performance outcomes were measured using data from school records. Data on attendance, scores on the state reading and math tests, promotion and retention, as well as high school dropout and graduation rates, were collected from these records.

In addition, the state provided information on reading and math scores for students across Kansas. These data allowed us to compare the gains made by Kansas City students with those made by students in the state as a whole.

There were significant improvements associated with FTF implementation for most of these student outcomes.

Attendance

Charts XI-1 and XI-2 present changes in attendance over the first three years of FTF implementation (and in the fourth year for students attending schools in the Wyandotte and Washington clusters). These charts show a steady increase in the likelihood that secondary students met district attendance standards over the course of FTF implementation. Further, students at all grade levels were less likely to fall below district attendance standards with each additional year of FTF implementation.

Although increases in students' attendance occurred at all school levels, the improvements were strongest at the middle and high school levels. FTF had little relationship with the likelihood of elementary students meeting district attendance standards, primarily because virtually all elementary students (99%) were already meeting the standards prior to FTF implementation.

For middle school students, however, the likelihood of meeting or exceeding district attendance standards increased nearly twofold after three years of implementation compared with before FTF began. That is, for every 100 middle school students who were meeting attendance standards before FTF, an additional 87 were attending school at this level after three years of implementation. Improvements in attendance were even greater among high school students. At this level, for every 100 students who were meeting district attendance standards before FTF, an additional 153 were meeting this standard after three years.

There were also decreases in the likelihood of students falling below district attendance standards. Among elementary and middle school students, the probability of falling below district attendance standards declined by about 80 percent. Although this is a statistically significant drop, at the elementary level it reflects a change in attendance for relatively few students.[22] Among high school students, the likelihood of poor attendance declined by nearly 60 percent; that is, for every 100 students who fell below district attendance standards at baseline, only 41 had poor attendance after three years of implementation.

In order to examine whether improvements in attendance were occurring for all groups, we also examined whether changes in attendance varied by socio-economic status and by ethnicity (see Charts XI-3 to XI-11). There were no significant differences at the elementary level between students who received free or reduced-price lunch and those students who did not receive this assistance during FTF implementation.

At the middle school level, students who received free or reduced-price lunch services were 116 percent more likely to have met district standards in buildings with three years of implementation compared with an increased likelihood of 75 percent among those not receiving these services. This

[22] At baseline, only 157 students were below district attendance standards; in Year 3, this number dropped to only six students.

means that for every 100 low socio-economic status (SES) middle school students who met district attendance standards at baseline, 116 additional students met these standards in buildings that implemented FTF for three years. Among high SES students, for every 100 who met attendance standards at baseline, an additional 75 met them after three years of implementation. Thus, improvement in attendance was greater among low SES students compared with high SES students.[23]

At the high school level, however, the attendance gap between students receiving free or reduced-price lunch and those not receiving assistance increased after three years of implementation. The likelihood of meeting attendance standards increased by 136 percent among high SES students, but only by 70 percent among low SES high school students.

Results for poor attendance showed a similar pattern. At the middle school level, the likelihood of falling below district attendance standards decreased by 83 percent for low SES students compared with a decrease of only 79 percent among high SES students. At the high school level, however, the probability of falling below district standards decreased by 59 percent among high SES students, but only by 37 percent for low SES students.

Looking at differences in attendance rates for different racial groups, we found that improvements in attendance after three years of implementation were greatest among Hispanic students at the middle school level. The likelihood of meeting district attendance standards increased 145 percent among Hispanic students, by 115 percent for white students and by 86 percent for African-American students. Thus, the gap between white and Hispanic youth decreased while the gap between white and African-American youth increased.

At the high school level, however, attendance improved most for African-American students (104% more likely to have met district standards) compared with white high school students (who were 81% more likely to have met standards) and Hispanic students (who were 43% more likely to have met district standards).

Chart XI-1

Increase in Likelihood of Attendance
Meeting/Exceeding District Standards

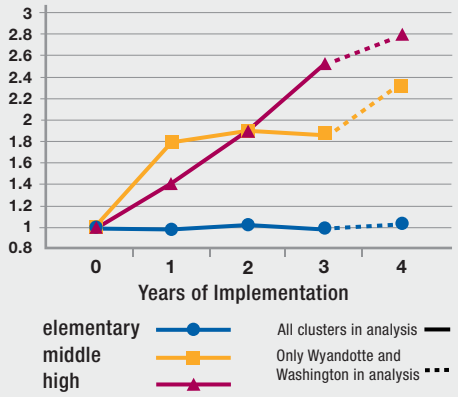


Chart XI-2

Decrease in Likelihood of Attendance
Falling Below District Standards

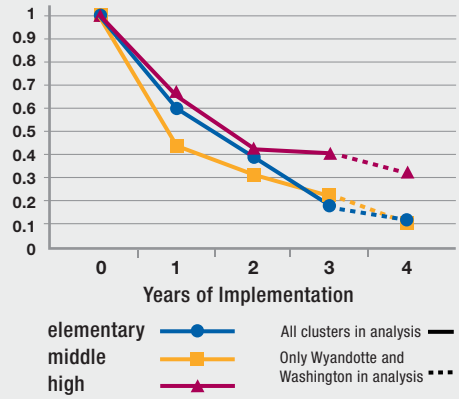
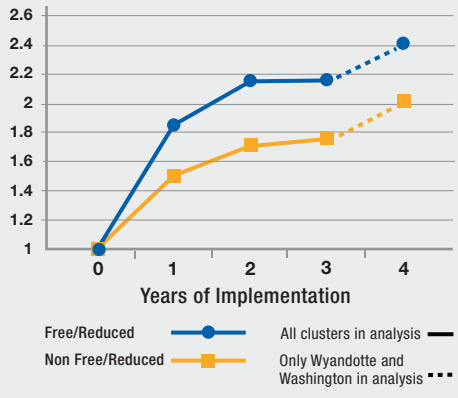


Chart XI-3

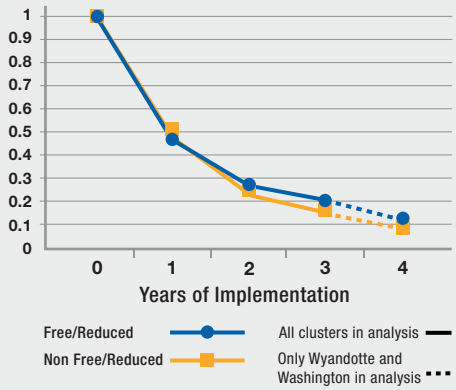
Increase in Likelihood of Good Attendance
Middle School



[23] Although the gap appears to be closing again after four years of implementation, data are only available for the Wyandotte and Washington clusters for this year. Whether the same pattern would be evident for the district as a whole cannot be determined until the remaining clusters complete their fourth year of implementation.

Chart XI-4

Decrease in Likelihood of Poor Attendance
Middle School



The results for poor attendance also show Hispanic youth in middle school were closing the gap between their white counterparts (a decrease in the likelihood of falling below standards of 91% among Hispanic youth compared with a 73% decrease among white students). The decrease in probability of having poor attendance was similar between African-American and white students in middle school.

At the high school level, however, African-American youth were closing the gap between themselves and white students (the likelihood of falling below district standards decreased 46% among African-American students compared with a 39% decrease among white students). The decrease among Hispanic high school students was similar to that observed among white high school students.

Chart XI-5

Increase in Likelihood of Good Attendance
High School

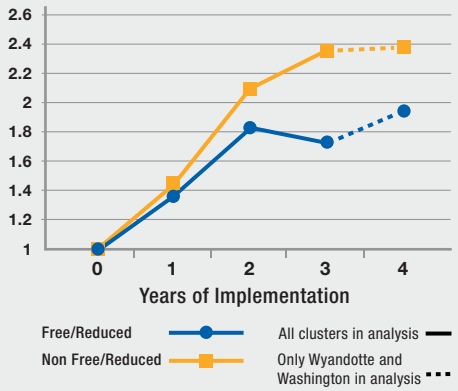


Chart XI-7

Increase in Likelihood of Good Attendance
Elementary

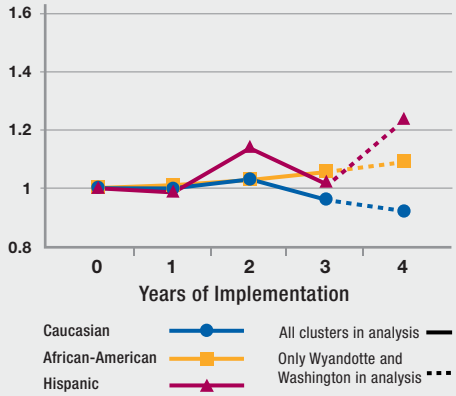


Chart XI-6

Decrease in Likelihood of Poor Attendance
High School

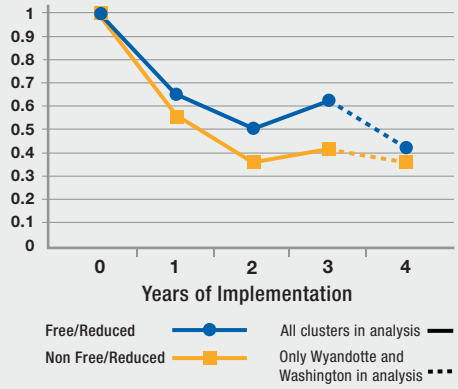


Chart XI-8

Increase in Likelihood of Good Attendance
Middle School

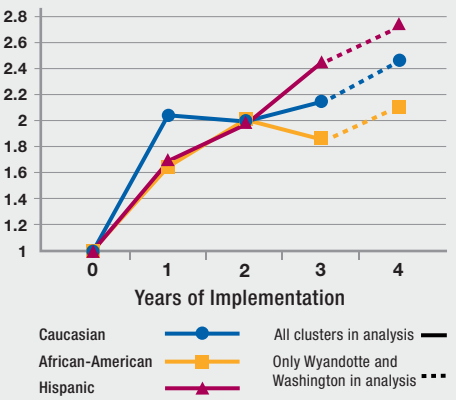


Chart XI-9

Decrease in Likelihood of Poor Attendance
Middle School

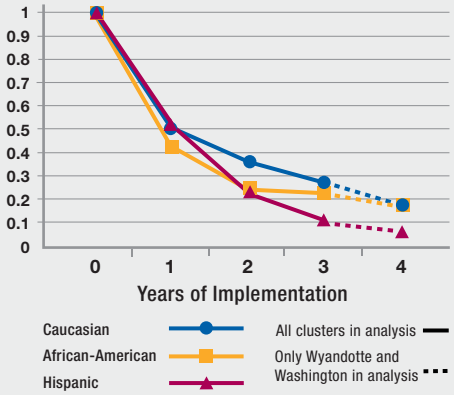


Chart XI-10

Increase in Likelihood of Good Attendance
High School

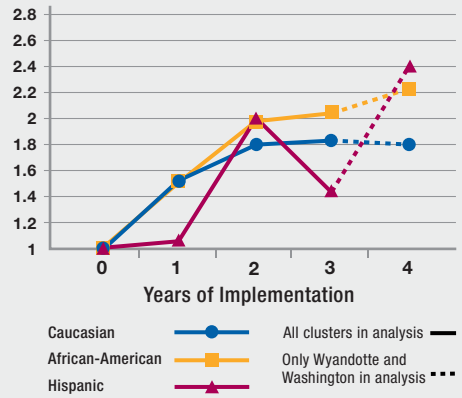
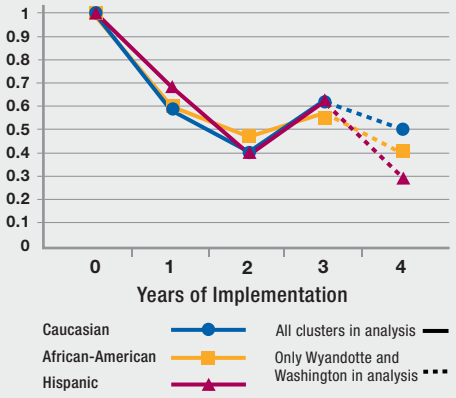


Chart XI-11

Decrease in Likelihood of Poor Attendance
High School



State Reading Test

Because Kansas changed its state reading and math assessments beginning with the 2000-2001 school year, we were only able to examine changes in test scores over three school years regardless of the number of years of implementation within each cluster. The results discussed below examine the likelihood of scoring at the proficient (or unsatisfactory) level in the 2002-2003 school year compared with 2000-2001. To examine the relationship between FTF implementation and student achievement, later test scores were compared with those in the first year of implementation (available only for schools in the Harmon and Schlagle clusters). Charts XI-12 and XI-13 present the changes in the likelihood of scoring proficient or unsatisfactory after each additional year of implementation.

In the first year of FTF implementation, 29 percent of elementary students, 29 percent of middle school students and 18 percent of high school students in Kansas City scored proficient or above on the state reading test. At the same time, between 35 percent (among middle school students) and 46 percent (among high school students) scored unsatisfactory or below on the state test. After three years of implementation, students at all levels were more likely to score proficient or above and less likely to score unsatisfactory on the test. Elementary and middle school students were about 1.6 times and high school students were about 1.4 times as likely to score proficient after three years of FTF.

Students were also significantly less likely to score unsatisfactory on the state reading test over time. Elementary and middle school students were about 50 percent less likely to score unsatisfactory and high school students were slightly less (about 4%) likely to score poorly on the reading test after three years of implementation.

To further explore whether the improvements in reading test scores are associated with the introduction of FTF into the KCK district, we compared trends in test scores for the district with trends at the state level. For this analysis, however, changes in test scores for each school year rather than after each additional year of FTF implementation were examined. During the initial year the current state test was administered (i.e., 2000-2001), the percentage of students attaining a proficient score was considerably higher for the state overall compared with Kansas City students. For example, at the

Chart XI-12

Increase in Likelihood of Scoring Proficient or Above: State Reading

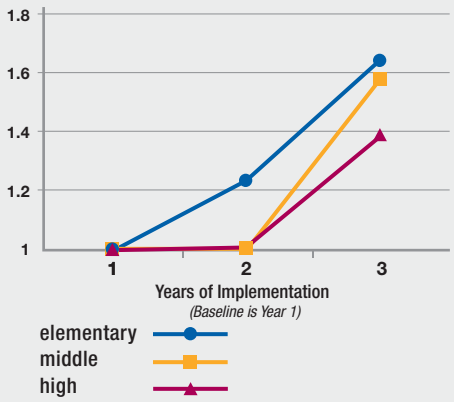


Chart XI-13

Decrease in Likelihood of Scoring Unsatisfactory: State Reading

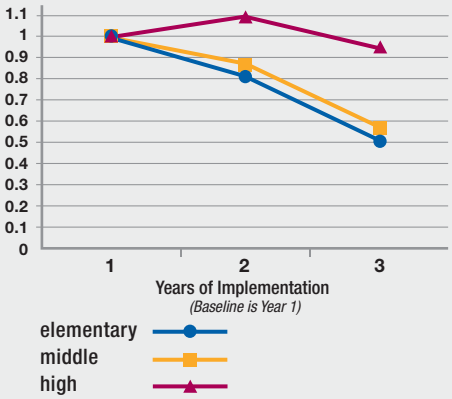
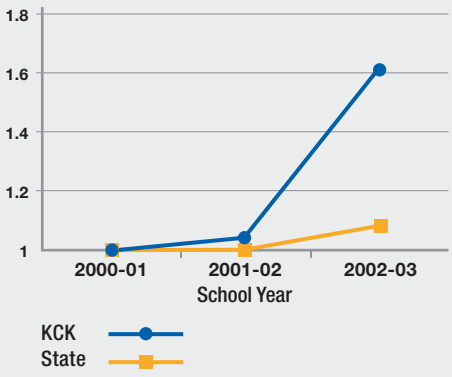


Chart XI-14

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01 Elementary Reading



elementary level, 66 percent of students in the state achieved a proficient score in reading compared with only 31 percent of Kansas City students. Similar gaps between the state and Kansas City were observed for middle and high school students.

Over the next three years, that gap narrowed, with greater increases in the likelihood of scoring at the proficient level occurring in Kansas City relative to increases for all other students across the state. For example, middle school students in Kansas City were 62 percent as likely to attain a proficient score on the reading test in the 2002-2003 school year compared with the 2000-2001 school year. In contrast, middle school students across the state were only about 7 percent more likely to score at the proficient level by Year 3 compared with baseline. (See Charts XI-14, 15, 16, 17, 18 and 19.)

Comparison of improvements in reading scores for different ethnic groups in Kansas City with those for students across the state showed that minority students in Kansas City were closing the gap with white students at a greater rate than in the remainder of the state. That is, improvements in reading test scores in Kansas City were greatest among ethnic minority groups. For example, African-American elementary students in Kansas City were nearly 70 percent more likely to score proficient in reading compared with only 12 percent of African-American students across the state. White students in elementary school in Kansas City were 18 percent more likely to score proficient, while those in the rest of the state were 6 percent more likely. (See Charts XI-20, 21 and 22.)

At the middle school level, Hispanic students in Kansas City were more than 100 percent more likely to attain a proficient score in 2002-2003 relative to baseline, whereas Hispanic students across the state were only about 40 percent more likely to score at this level (a difference of 74 percentage points). In contrast, white middle school students in Kansas City were 27 percent more likely to score proficient compared with white students elsewhere, who were only 4 percent more likely to score well (a difference of 23 percentage points).

There was a similar pattern with regard to decreases in the likelihood of scoring unsatisfactory or below on the state reading test at the middle and high school levels.[24] For all ethnic groups, the decline in the probability of scoring poorly on the test was greater among Kansas City students than declines for similar

[24] At the elementary school level, the likelihood of scoring unsatisfactory decreased by a larger percentage among minority students for the state overall than in Kansas City.

students across the state. For example, African-American middle school students in Kansas City were 60 percent less likely to score unsatisfactory by 2002-2003 compared with others across the state, who were 15 percent less likely to have poor reading scores (see Charts XI-23, 24 and 25).

The greater increases in the likelihood of scoring at the proficient level combined with the greater declines in the likelihood of scoring at the unsatisfactory level in Kansas City has resulted in a narrowing of the gap between Kansas City students and students across the state. In most cases, these increases have been greater among ethnic minority groups, which have also resulted in narrowing the achievement gap between white and minority students in Kansas City at a faster pace than observed for the remainder of the state.

Turning to an examination of socio-economic status (SES), we found that at the elementary and high school levels, low SES students were closing the gap with their high SES counterparts at a higher rate in Kansas City compared with the rest of the state (not shown). Among elementary students, the likelihood of scoring at the proficient level increased by 80 percent among low SES students in Kansas City compared with a 17 percent increase for the rest of the state (a 68 percentage point advantage). Whereas, among high SES students in elementary school, the corresponding likelihood increases were 24 percent (Kansas City) and 6 percent (state) (only an 18 percentage point difference).

Results were similar at the high school level, where students receiving free or reduced-price lunch were closing the gap between their counterparts who did not receive this assistance at a faster rate than students in the rest of the state. At the middle school level, however, there were no significant differences between high and low SES students in Kansas City and other high and low SES students across the state.

Chart XI-15

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01
Elementary Reading

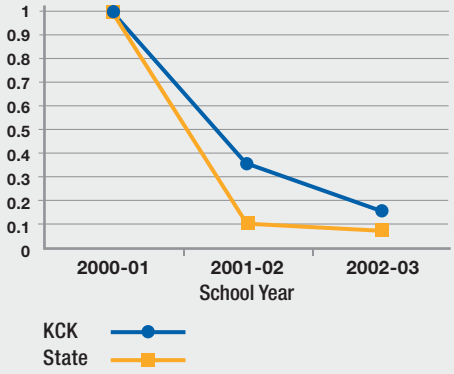


Chart XI-16

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01
Middle School Reading

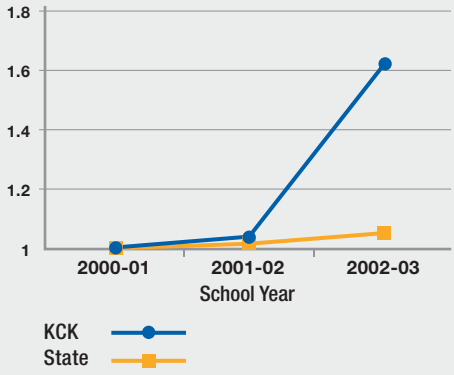


Chart XI-17

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01
Middle School Reading

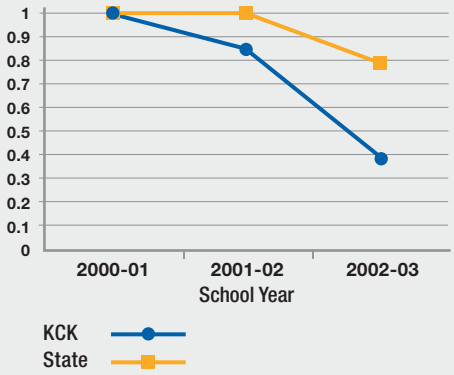


Chart XI-18

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01
High School Reading

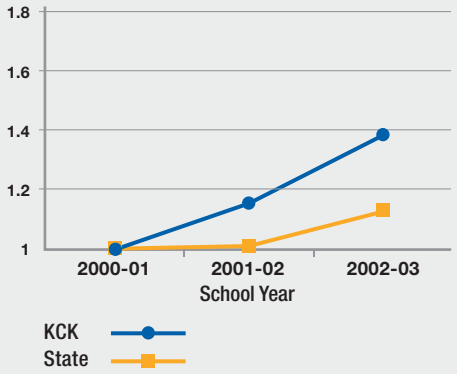


Chart XI-19

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01
High School Reading

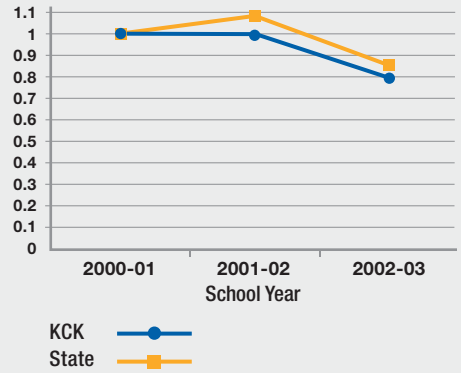


Chart XI-20

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01 to School Year 02-03
Elementary Reading

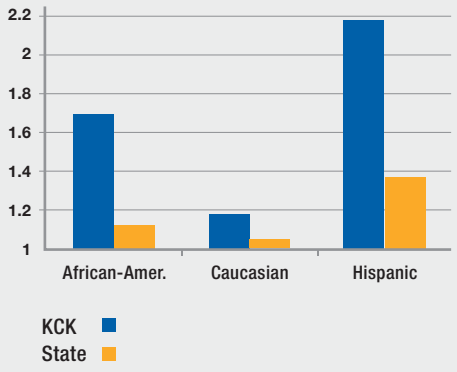


Chart XI-21

Increase in Likelihood of Scoring Proficient or Higher from School Year 00-01 to School Year 02-03
Middle School Reading

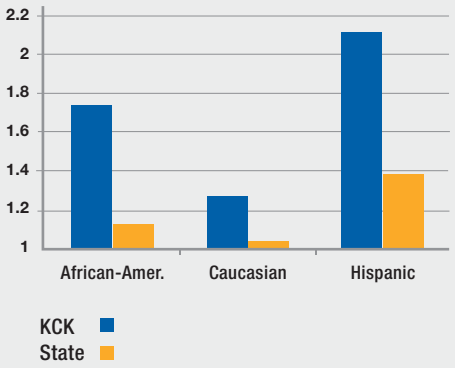


Chart XI-22

Increase in Likelihood of Scoring Proficient or Higher from School Year 00-01 to School Year 02-03
High School Reading

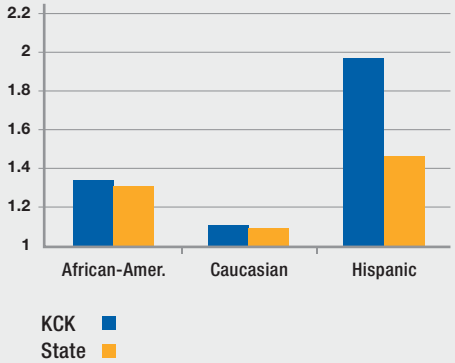


Chart XI-23

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01 to School Year 02-03
Elementary Reading

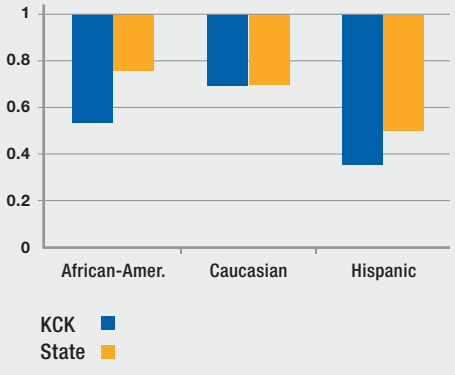


Chart XI-24

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01 to School Year 02-03 Middle School Reading

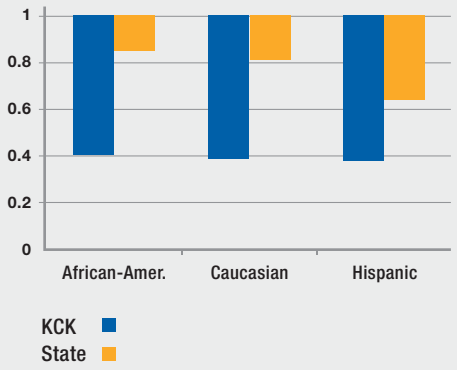
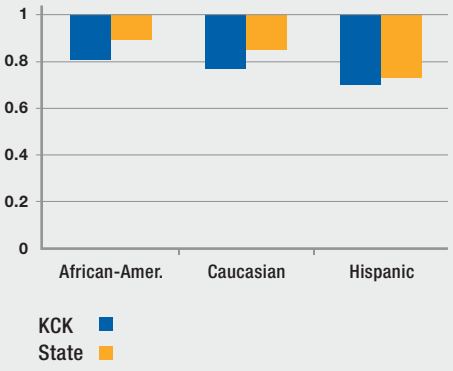


Chart XI-25

Decrease in Likelihood of Scoring Unsatisfactory from School Year 00-01 to School Year 02-03 High School Reading



State Math Test

Similar to the findings for the state reading test, we found that students' scores on the state math test also improved over the course of FTF implementation. After one year of FTF implementation, 33 percent of elementary students, 12 percent of middle school students and 5 percent of high school students scored proficient or above on the state math test. Looking at the three-year trends (shown in Chart XI-26), there was a steady increase in the probability of elementary students scoring at the proficient level in math. Although middle and high school students were also more likely to score at the proficient level after additional years of FTF implementation, scores did not steadily increase at these levels. After three years of implementation, elementary and middle school students were about 1.5 times more likely to attain proficient scores, while high school students were nearly twice as likely to score at the proficient level.

There were also significant declines in the probability of scoring at the unsatisfactory level on the state math test. After the first year of FTF implementation, 35 percent of elementary, 63 percent of middle school and 74 percent of high school students scored unsatisfactory or below on the test. After three years of implementation, elementary students were about 40 percent less likely to score poorly in math (see Chart XI-27). Middle and high school

Chart XI-26

Increase in Likelihood of Scoring Proficient or Above: State Math

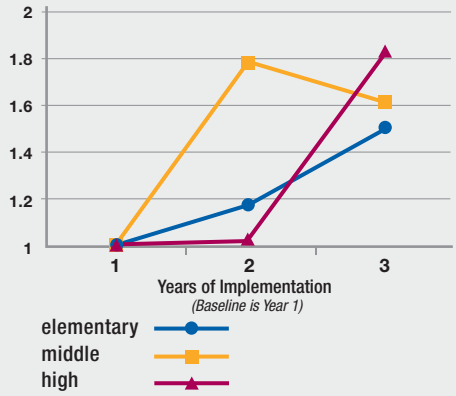


Chart XI-27

Decrease in Likelihood of Scoring Unsatisfactory: State Math

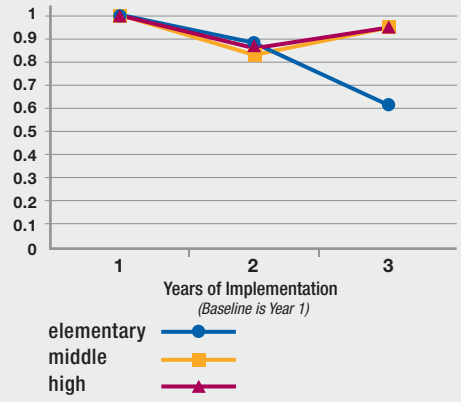


Chart XI-28

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01
Elementary Math

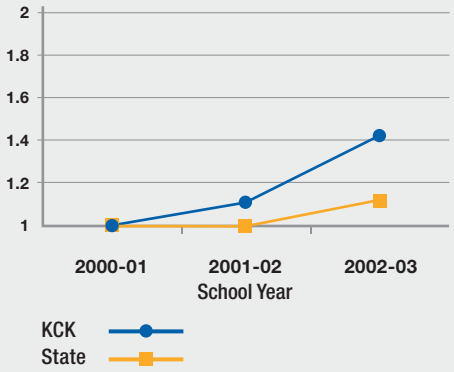


Chart XI-29

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01
Middle School Math

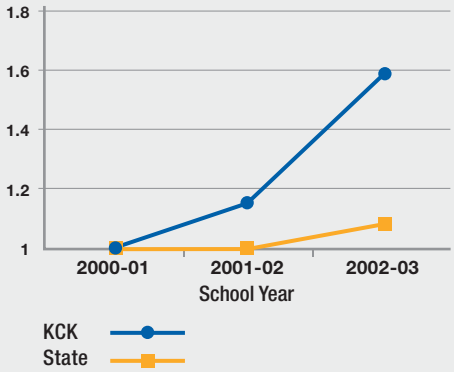
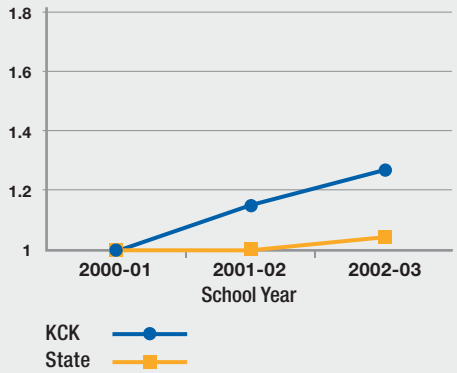


Chart XI-30

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01
High School Math



students were about 15 percent less likely to score in the unsatisfactory range of the test. At the elementary level, for example, for every 100 students who scored unsatisfactory in the first year, only 63 scored unsatisfactory after three years of implementation.

Comparing changes in math test scores in Kansas City with other students across the state, we observed (similar to the results for the reading test) that students in Kansas City were much less likely in the 2000-2001 school year to score at the proficient level and more likely to score unsatisfactory compared with students in the state, overall. For example, 69 percent of elementary students across the state attained proficient scores in math during the first year of the test compared with only 31 percent of Kansas City elementary students.

By the third year of the test (the 2002-2003 school year), however, the gap between Kansas City students and other students across the state had narrowed, especially at the elementary and middle school levels (see Charts XI-28, 29 and 30).

This narrowing of the gap was particularly evident among minority students. For example, African-American elementary students were 60 percent more likely to attain a proficient score on the math test by Year 3 compared with other African-American students across the state, who were 17 percent more likely to attain a math score at the proficient level (a difference of 43 percentage points). Increases for white students at this level were 19 percent in Kansas City and 8 percent for the rest of the state (a difference of only 11 percentage points).

Similarly, Hispanic students in Kansas City middle schools were 64 percent more likely to score at the proficient level compared with other Hispanic students across the state, who were 13 percent more likely to score well on the math test in Year 3 relative to baseline. Although the likelihood of scoring proficient for middle school white students in Kansas City also increased at a greater rate than for other students in the state (28% compared with 6%), the differential between Kansas City and the state is smaller than that for Hispanic students (a 22 percentage point difference among white students compared with a 51 percentage point difference among Hispanic students). (See Charts XI-31, 32 and 33.)

Regarding the likelihood of scoring unsatisfactory on the state math tests, differences between Kansas

City students and those across the state were less prevalent. At the elementary school level, students in both Kansas City and the remainder of the state were about one-third less likely to score unsatisfactory in the third year of the test. Kansas City middle school students were about 20 percent less likely to score poorly in math compared with a 15 percent decrease in the likelihood of scoring poorly for the rest of the state. There was little change in the likelihood of scoring unsatisfactory in math at the high school level, in either Kansas City or across the state as a whole. (See Charts XI-34, 35 and 36.)

Turning to the likelihood of scoring unsatisfactory on the math test for different racial groups, there were different results across school levels. Among elementary students, ethnic minority groups had greater decreases in the likelihood of scoring unsatisfactory compared with other minority students across the state; but among white students, decreases were greater for the rest of the state than observed among Kansas City students (see Chart XI-37). Thus, the gap between white and minority students was closing in Kansas City, while it was increasing for other students in the state.

In middle school, however, Kansas City students in all ethnic groups had greater decreases in the probability of scoring poorly in math compared with similar students across the state. At the high school level, there was virtually no change in the likelihood of scoring at the unsatisfactory level – for any ethnic group in either Kansas City or across the state. (See Charts XI-38 and 39.)

With regard to SES status, students in Kansas City were comparable to their counterparts across the state on the math test (not shown). In general, changes in the probability of scoring at either the proficient or unsatisfactory level were similar in Kansas City and across the state.

Chart XI-31

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01 to School Year 02-03
Elementary Math

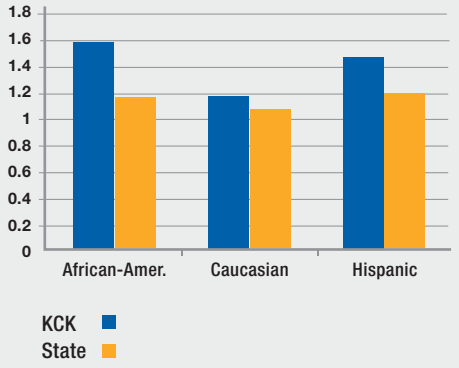


Chart XI-32

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01 to School Year 02-03
Middle School Math

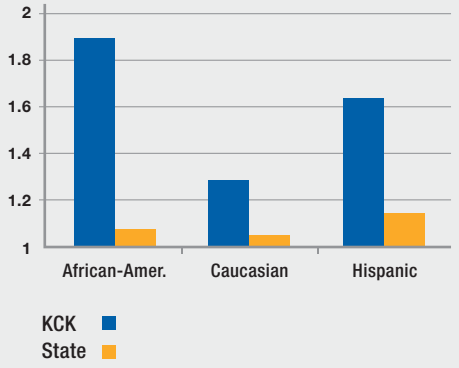


Chart XI-33

Increase in Likelihood of Scoring Proficient or Above from School Year 00-01 to School Year 02-03
High School Math

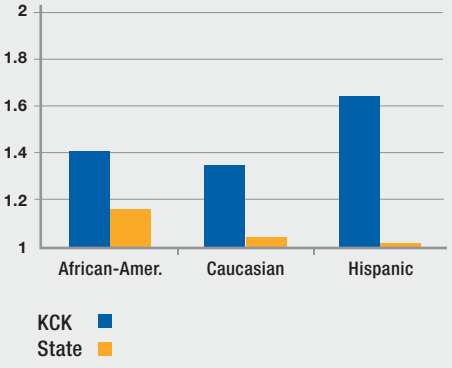


Chart XI-34

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01
Elementary Math

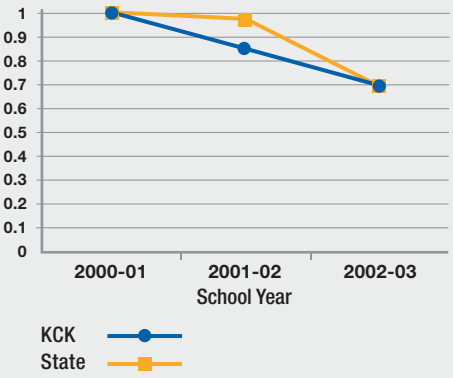


Chart XI-35

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01
Middle School Math

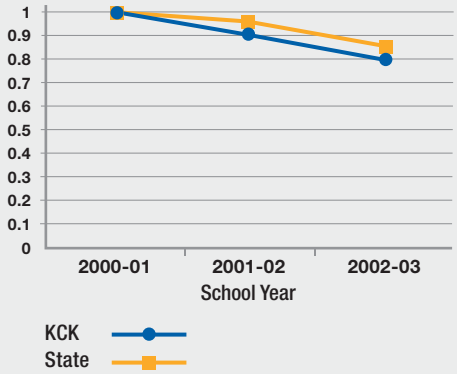


Chart XI-36

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01
High School Math

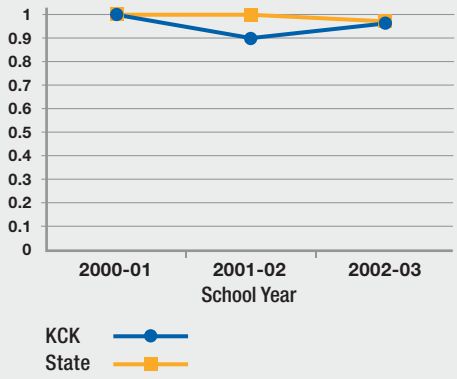


Chart XI-37

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01 to School Year 02-03
Elementary Math

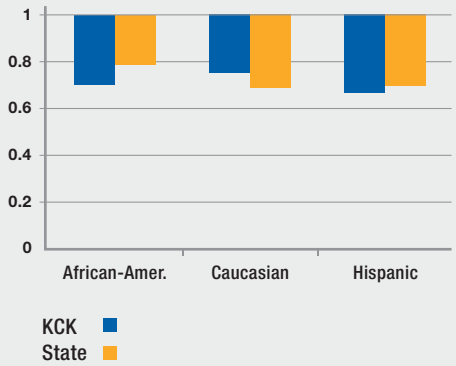


Chart XI-38

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01 to School Year 02-03
Middle School Math

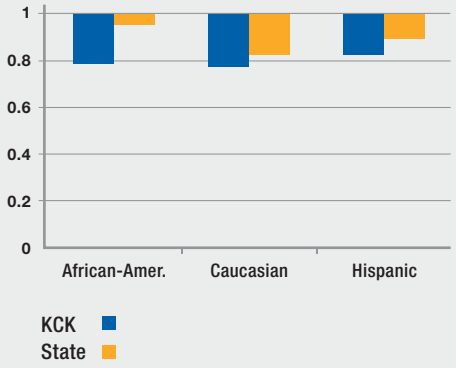
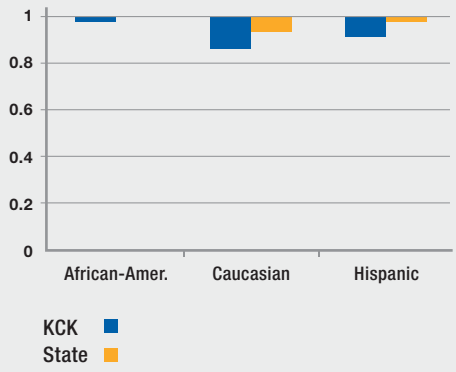


Chart XI-39

Decrease in Likelihood of Scoring Unsatisfactory
from School Year 00-01 to School Year 02-03
High School Math



Graduation

The district was able to provide accurate records of graduation rates starting in 1997-1998. Because of this, we could only look at three graduation cohorts beginning with the graduating class of 2000-2001, which corresponds to different years of FTF implementation for each of the four high schools. As a result, we examined changes in the probability of each cohort graduating separately by cluster. For each high school, we compared the likelihood of graduating for the 2002-2003 graduation class relative to the 2000-2001 graduation class. Thus, for Wyandotte high school, the comparison is between the third and fifth years of implementation; for Washington high school, the likelihood of graduating after four years of FTF is compared with graduation rates following the second year of implementation; for Harmon and Schlagle, the probability of graduating after three years of implementation is compared with the graduation rates in the first year of FTF implementation in those high schools (see Chart XI-40).

Graduation rates for the class of 2000-2001 ranged between 64 and 69 percent across the four high schools. There were small, but significant, changes in the likelihood of graduating across the three cohorts examined for each cluster. In general, students in the class of 2002-2003 were between 20 percent (in Wyandotte) and 27 percent (in Washington) more likely to graduate compared with graduation rates for the class of 2000-2001 –meaning that graduation rates would increase to more than 80 to 85 percent, holding other factors (such as SES and ethnicity) constant.

Chart XI-40

Increase in Likelihood of Graduating Compared to High School Class 2000-01

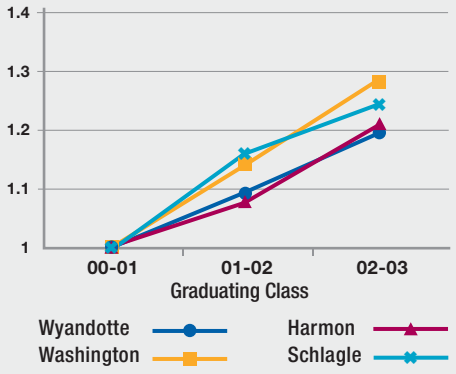
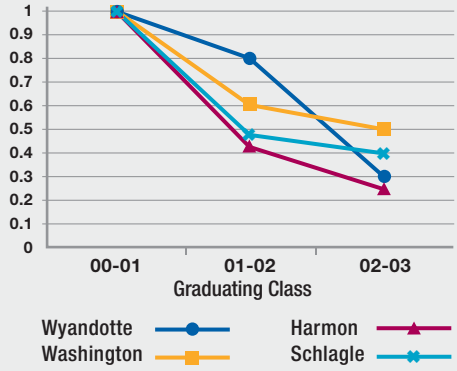


Chart XI-41

Decrease in High School Dropout Compared to Graduating Class 2000-01



Dropping Out

The district was able to provide accurate records on high school dropout rates beginning with the 1998-1999 school year. Thus, for this analysis, we also examined changes for each of the four high schools separately rather than in the aggregate. For each high school, we compared the likelihood of dropping out of high school for the 2002-2003 graduation class relative to the 2000-2001 graduation class. Thus, for Wyandotte high school, the comparison is between the third and fifth years of implementation; for Washington high school, the likelihood of dropping out after four years of FTF is compared with dropout rates following the second year of implementation; for Harmon and Schlagle, the probability

of dropping out after three years of implementation is compared with the dropout rates in the first year of FTF implementation in those high schools (see Chart XI-41).

The probability of dropping out of high school declined significantly in each of the four high schools over the course of FTF implementation. In Wyandotte High School, for example, at baseline, about 18 percent of the students dropped out before graduating. After four years of FTF implementation, students were about 70 percent less likely to drop out of school. That means that for every 100 students who dropped out in the baseline year, only 30 would have dropped out after four years of implementation.

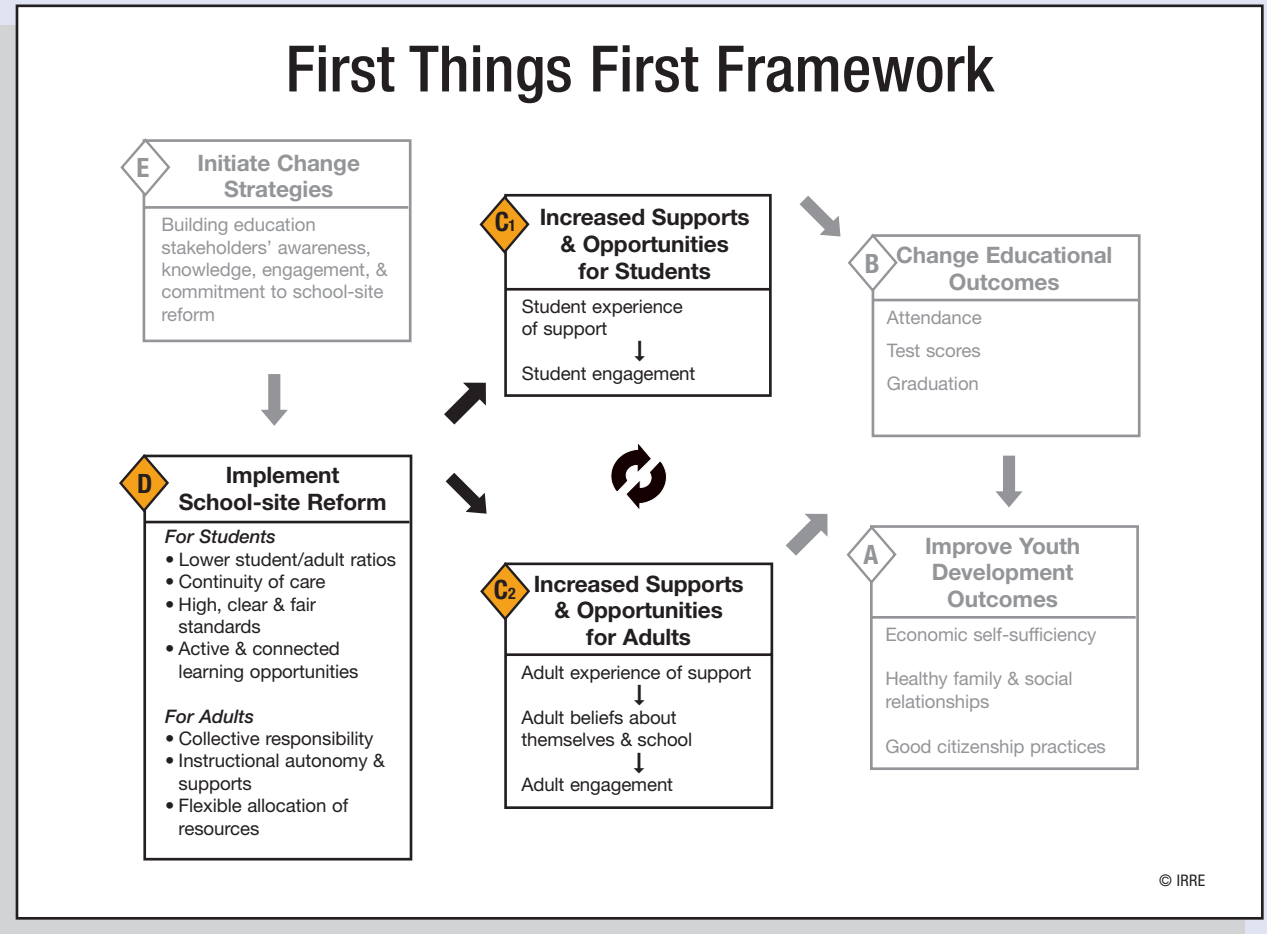
EXPLORING PATHWAYS TO CHANGE

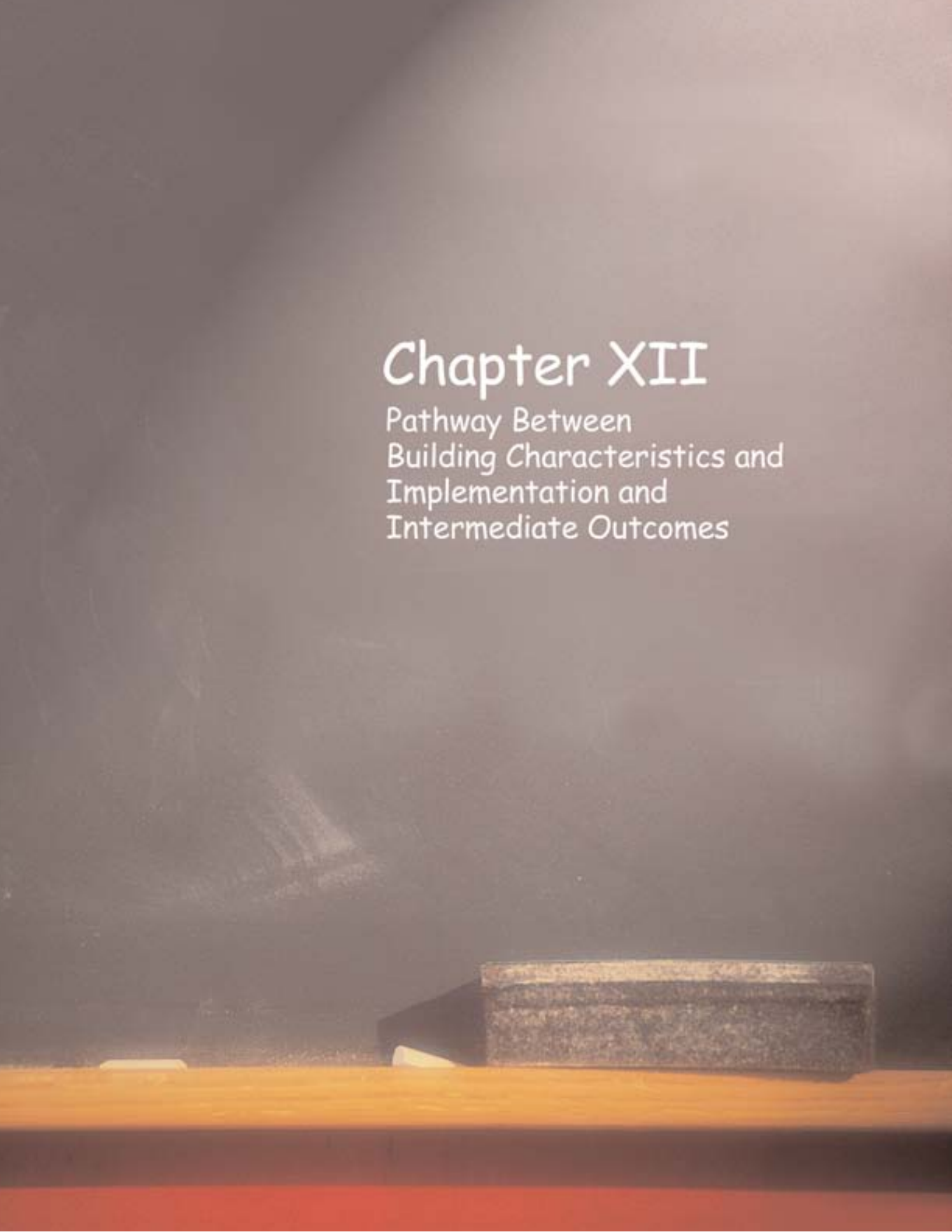
The FTF model predicts that if a school district creates the necessary conditions and supports for change and implements the critical features of school-site reform, then student and staff experiences of support (i.e., the intermediate outcomes) will increase. In turn, the model posits that if the intermediate outcomes are increased then long-term outcomes, such as attendance and test scores, will improve. Chapters XII to XIV explore whether these hypotheses held in the Kansas City, Kansas, school district. Each chapter contains information about the statistically significant results (in the predicted direction or not), along with a figure that exhibits the relative size of the relationship between the outcomes. Chapter XIII focuses on the relationship between implementation and intermediate outcomes; Chapter XIV on the relationship between intermediate and long-term outcomes in the FTF model.

TESTING THE PATHWAYS BETWEEN OUTCOMES

QUESTION 4: What characteristics of leaders and staff in buildings seem to facilitate or impede implementation?

QUESTION 5: Are improved classroom structures and instruction associated with better intermediate outcomes (student relationships with teachers and engagement in school)?





Chapter XII

Pathway Between Building Characteristics and Implementation and Intermediate Outcomes

MAJOR FINDINGS

- *Buildings observed to have strong relationships among and between staff and students also had more classrooms with optimal ratios; more highly engaged staff; more staff feeling supported by colleagues and system leaders; and more students reporting high classroom standards, quality relationships with their teachers and high engagement in schools.*
- *Buildings with strong principal and SLC leadership were significantly more likely to have implemented high-quality instruction and standards, and to have more students and staff feeling highly engaged and supported.*
- *High-quality structures (i.e., small class sizes and consistent exposure to SLCs) are more likely to be found in buildings with a high-quality principal, strong SLC leadership and participative decision-making structures.*
- *Buildings with high-quality professional development are more likely to use small grouping strategies.*

This chapter focuses first on whether characteristics of leaders and staff in buildings are related to the strength of implementation. In the framework picture, the building characteristics would be represented by a box between the E box (initiative strategies) and D box (critical feature implementation). In other words, some of the capacity-building strategies used by initiative leaders were directed at specific critical features (e.g., ratios); where others were intended to increase the supports needed in buildings to implement critical features (e.g., principal leadership, staff professional development).

This chapter also includes a section on the relationship between these same building characteristics and the intermediate outcomes (student engagement and support, staff engagement and support). Since the original reform framework did not explicitly include these building characteristics and processes, we conducted exploratory analyses examining the effects on both implementation and the intermediate outcomes later in the pathway.

BUILDING CHARACTERISTICS MEASURED

The research team used a qualitative approach to observe and track the processes occurring within the buildings as they planned and implemented the reform. Over the course of five years, the schools in the Wyandotte and Washington clusters were each visited several times a year for individual interviews, focus groups, observations of team meetings and professional development sessions, and "shadowing" students as they moved through a school day. In this process, the qualitative team identified seven characteristics of leaders and staff in buildings that appeared to make a difference in either facilitating or impeding implementation. For each characteristic, the team developed a set of indicators that provided a way to rate the degree to which a school exemplified each. These characteristics and their indicators are:

Principal Leadership Abilities
- Staff believe the principal cares about them
- The principal is highly visible in classrooms and hallways and "pitches in" with instruction
- Staff believe the principal knows what is going on
- Communication between principal and staff is open and direct

School Improvement Facilitator Leadership
- The SIF has the trust of the principal
- The SIF demonstrates competence in instructional leadership
- Staff turn to the SIF with instructional questions

SLC Coordinator Leadership (Teacher Leadership)
- The coordinator position in a SLC is stable (not rotated among team members)
- The coordinator is well-accepted by the other teachers
- The coordinator demonstrates effective facilitation skills

Decision-Making Processes
- Staff have information to make decisions
- Staff believe their decisions are respected
- The group has a mechanism for free discussion
- The group has a mechanism to resolve differences
- SLC teams meet regularly

Professional Development Activities
- Participants perceive the topic as relevant to them
- Participants are actively engaged in the session
- Training facilitator models good practices
- Participants have hands-on, small group, activities
- Participants have concrete ways to assess the effect of the new practice
- Participants have concrete assignments to practice the skills presented
- Follow-up sessions provide opportunity for feedback

Peer (Staff) Relationships
- There is a system for incorporating and mentoring new staff

- Social groups are neither rigid nor fueled by negative feelings
- Communication is open and based on mutual respect
- The majority of conversations are constructive rather than complaints

Staff-Student Relationships

- Students can identify at least one adult they would turn to
- Staff demonstrate knowledge of students' lives outside of school.
- Staff avoid stereotyping groups of students
- Staff avoid blame statements or assumptions
- Staff are willing to listen and respond to student concerns
- Staff demonstrate a greater willingness to handle discipline

The qualitative research team created a rating scale to be used to assess each indicator on a scale of 1 to 5. The rubric for the scale is:

1 = School is a negative example of this indicator.
2 = School might be "at risk" with respect to this indicator.
3 = Neutral: School is neither a positive nor a negative example of this indicator.
4 = School displays this indicator fairly frequently.
5 = School displays this indicator consistently.

After rating the schools on each indicator, the quantitative team divided the ratings into three groups: high, medium and low levels for each factor. Analyses were then conducted to assess the association between the outcome variables and the three categories of factor scores (high, medium and low).[25]

The relationships between these characteristics and implementation of structural critical features, instructional critical features, and intermediate outcomes are explored in the remainder of this chapter.

PATHWAY BETWEEN BUILDING CHARACTERISTICS AND IMPLEMENTATION OF STRUCTURE

Observed Classroom Ratios

Figure XII-1 shows the relationships among the building factors and the proportion of classrooms that were optimal (less than 15:1 student-adult ratios) or high risk (more than 21:1 for elementary and 23:1 for secondary). High-quality relationships between staff and students, professional development, and principal leadership were all relatively strongly related to the likelihood of observing classrooms of 15 or fewer students. Buildings with high proportions of classrooms with 15 or fewer students were nine times more likely to have positive staff-student relationships. Schools with high-quality leadership were two and a half times more likely to have a high proportion of classrooms with 15 or fewer students, and buildings with high-quality professional development were three and a half times more likely. High-quality SLC leadership, participative decision-making and positive peer relationships were also significantly related to an increased likelihood of observing small classrooms, but the link was much weaker.

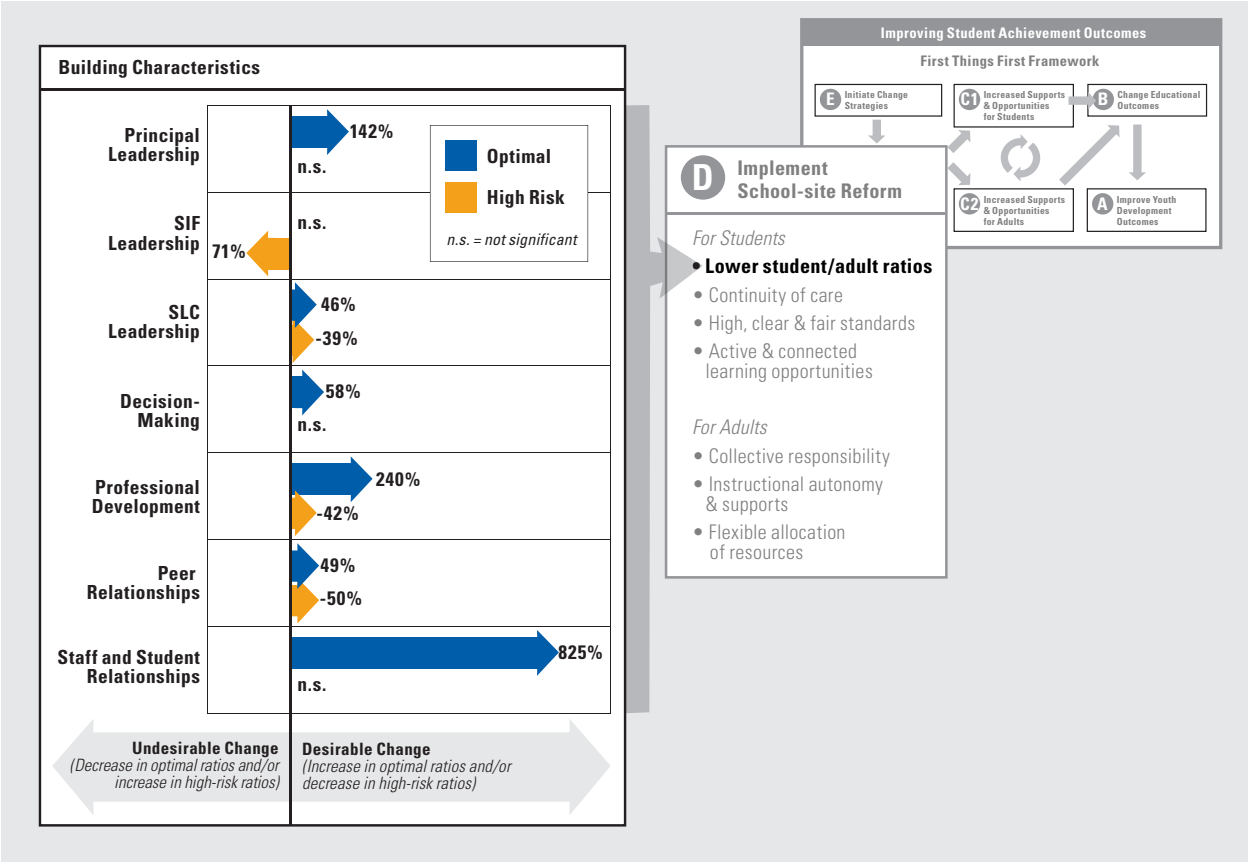
High-quality SLC leadership, professional development and positive staff relationships were also significantly related to decreases in the likelihood of observing large classes (39%, 42% and 50%, respectively). The only relationship not in the expected direction was high-quality SIF leadership, which was associated with a 71 percent increase in the likelihood of observing large classrooms.

SLC Stability

Figure XII-2 shows the relationships among the building factors and the number of students who were exposed to a SLC at least two years or more. Students in buildings with high-quality principal leadership were nearly six and a half times more likely to experience SLCs for at least two years. Similarly,

[25] Chi-Square analyses were performed to identify which of these correlations reached significance at $p < .05$ or less.

Figure XII-1: **Changes in the Percentage of Classrooms with Optimal (less than 15:1) and High-Risk (more than 21:1) Student-Adult Ratios Based on Classroom Observation**

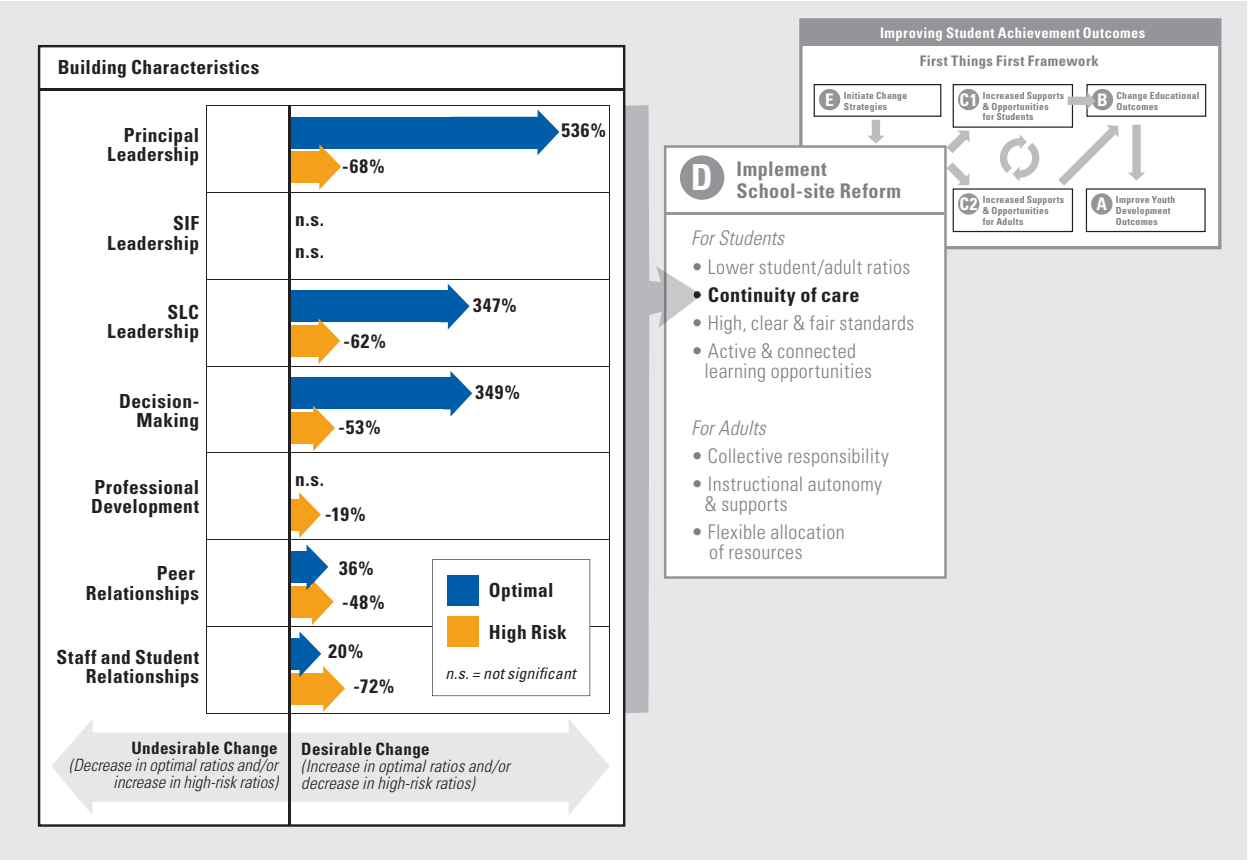


students in buildings with high-quality SLC leadership and participative decision-making processes were four and a half times more likely to experience SLCs for two or more years. Students in buildings with positive staff relationships and positive relationships between teachers and students were more likely to experience at least two years of SLC membership (36% and 20%, respectively). Students in buildings with high-quality principal leadership, SLC leadership, participative decision-making, and positive staff and staff-student relationships were substantially less likely to experience SLCs for less than two years.

Relationships between the building factors and the observational data for the use of small grouping structures, active learning, and classroom ratios appear to be weaker than the relationships between the building factors and student and staff variables. This may be an artifact of the methodology; that is, observations were taken at only one point in time for randomly selected classrooms; whereas, the staff and student variables were assessed through surveys that tapped the respondents' more general experiences across time and classrooms.

Closer inspection of the data suggest that significant, but non-linear results (not discussed), may reflect a "dip" in observational data during the 2001-2002 school year, when lower levels of all four variables were seen in comparison with the 2000-2001 and 2002-2003 school years. The SLC stability measure was not derived from observations but from district records, and therefore reflects a more general experience. The fact that SLC stability was not strongly related to professional development but was related to all the other leadership and relationship factors also makes sense, as placement decisions about students would conceivably not be affected by professional in-service training.

Figure XII-2: **Changes in the Percentage of Students Who Were Involved in a SLC for Two or More Years**



PATHWAY BETWEEN BUILDING CHARACTERISTICS AND IMPLEMENTATION OF INSTRUCTION

Student Report of Standards

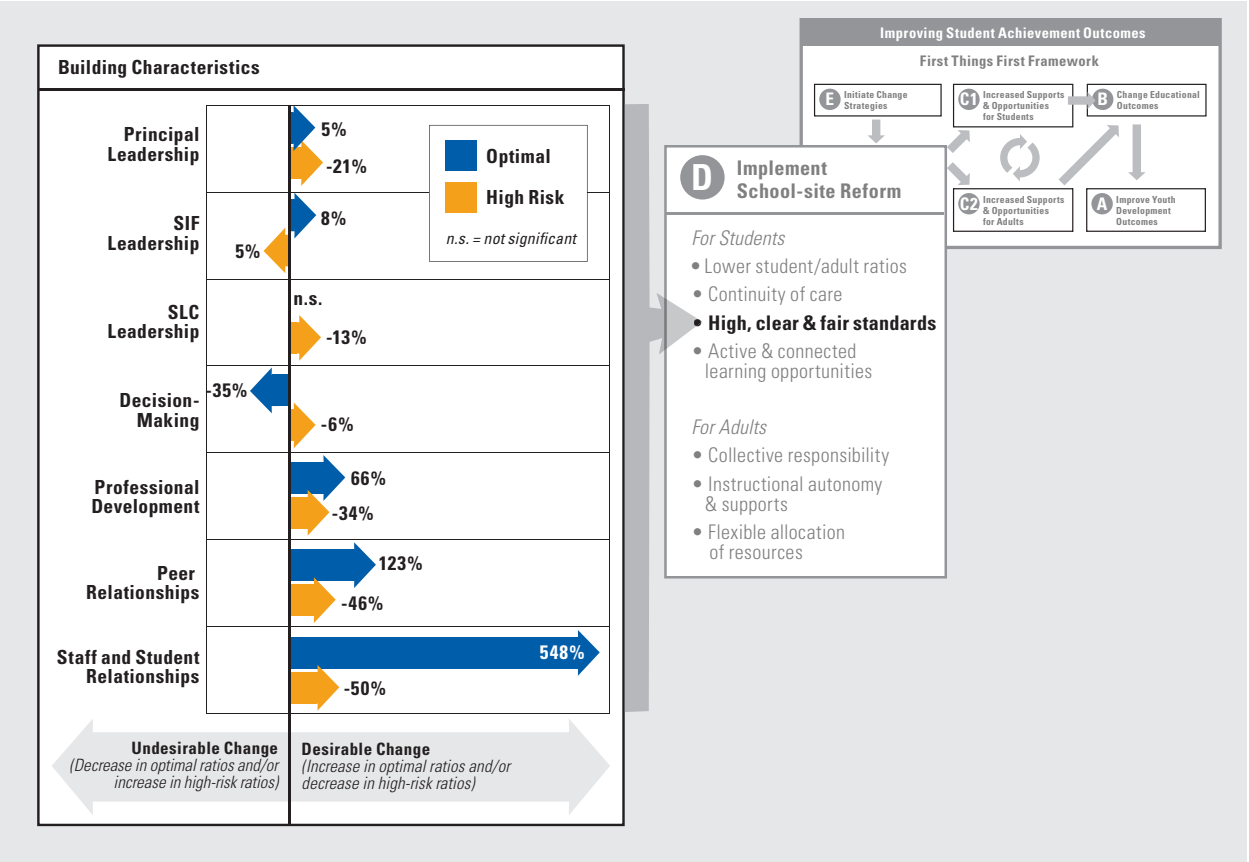
Figure XII-3 shows the relationships among the building factors and student experiences of academic and behavioral standards in the school. Similar to experiences of instruction, high-quality relationships were the strongest predictors of academic and behavioral standards. Students in buildings with high-quality staff relationships were 123 percent more likely to report experiencing consistently high levels of academic and behavioral standards. Students in buildings where there were strong relationships between students and staff were 548 percent more likely to experience consistently high levels of academic and behavioral standards.

Leadership was also significantly related to students' experiences of academic and behavioral standards. However, these effects were substantially smaller than the relationship effects and were strongest for reducing the likelihood that students would experience low levels of consistent academic and behavioral standards. For example, students in buildings with high-quality principal leadership and high-quality SLC leadership were 21 percent and 13 percent less likely to report low levels of academic and behavioral standards, respectively. SIF leadership was again negatively related to students' experiences of academic and behavioral standards. Students in buildings with high-quality professional development were 66 percent more likely to report high levels of consistent academic and behavioral standards, and 34 percent less likely to report low levels. Conversely, schools with more participative decision-making were 35 percent less likely to have students who experienced high standards.

Student Report of Active Learning

Figure XII-4 shows the relationships of leadership and building characteristics with students' experiences of instruction. Good relationships appear to be the strongest predictor of high-quality of

Figure XII-3: **Changes in the Percentage of Students Experiencing Optimal or High-Risk Levels of Academic and Behavioral Standards (Student Survey)**



instruction. Students were 37 percent more likely to report experiencing active learning strategies when staff relationships were high quality; and they were 28 percent more likely to do so when staff and students had good relationships.

Unsurprisingly, leadership is also important. Schools with strong SLC leadership were 24 percent more likely to have students reporting active learning opportunities and 23 percent less likely to experience passive learning strategies. Students in buildings with high-quality principal leadership and higher levels of participative decision-making were also significantly less likely to report experiencing passive learning strategies. However, the relationships between students' experiences of instruction and SIF leadership are negative, and buildings with high-quality principal leadership are 12 percent less likely to have students who report experiencing active learning in the classroom. The findings suggest that leadership quality is more likely related to decreasing students' experiences of passive learning strategies, but is mixed with respect to increasing their experiences of active learning opportunities.

Observed Active Learning

Figure XII-5 shows the relationships among the building factors and the type of instruction observed in classrooms. No significant relationships in the expected direction were found between observed implementation factors and observations of active learning. Participative decision-making and high-quality professional development were associated with small decreases in the likelihood of observing active learning strategies in the classrooms across buildings (32% and 37%, respectively).

Observed Grouping Structures

Figure XII-6 shows the relationships among the building factors and the frequency of the use of

Figure XII-4: **Changes in the Percentage of Students Experiencing Active or Passive Instruction (Student Survey)**

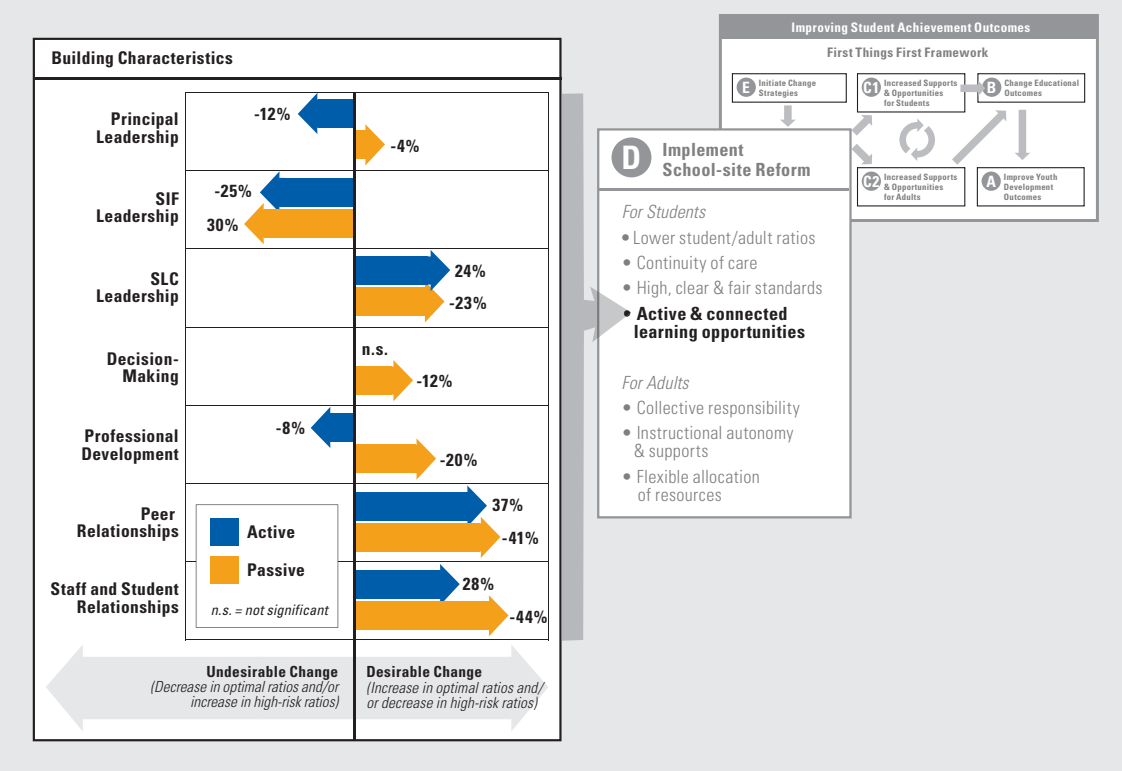
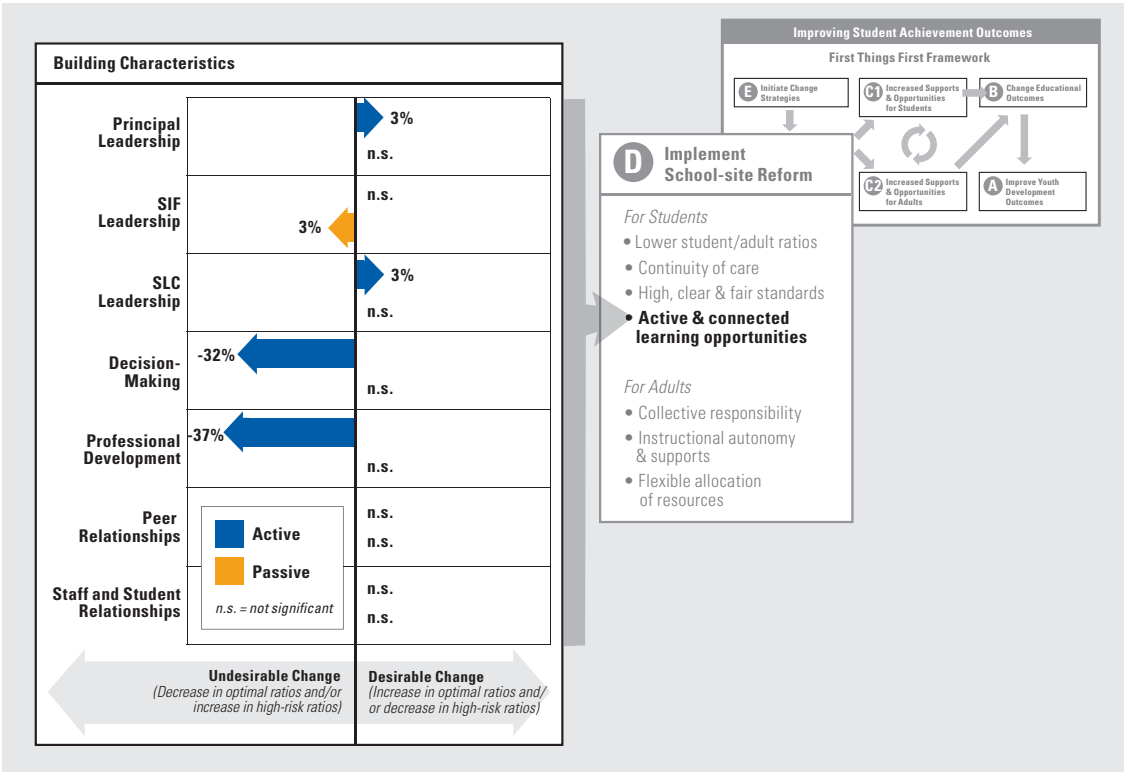


Figure XII-5: **Changes in the Percentage of Classrooms Exhibiting Active and Passive Learning Strategies (Classoom Observation)**



small grouping strategies observed in classrooms. High-quality professional development was most strongly associated with the likelihood of observing small grouping strategies in the classroom (278% increase in likelihood), and with a decrease in the likelihood of observing large or whole group instruction (61% decrease in likelihood). High-quality SLC leadership was also associated with a 48 percent increase in the likelihood of observing small grouping structures, and a similar decrease in the likelihood of observing large or whole group instruction. High-quality relationships among staff and students were associated with a 45 percent reduction in the likelihood of observing large and whole group instruction. High-quality principal leadership and relationships among staff were negatively associated with the likelihood of observing small grouping structures, an opposite effect than what was hypothesized.

PATHWAY BETWEEN BUILDING CHARACTERISTICS AND INTERMEDIATE OUTCOMES

Student Engagement

Figure XII-7 shows the relationships among the building factors and student engagement. Again, relationships and professional development were relatively strongly related to student engagement. Students in buildings with more positive relationships were between 147 percent and 289 percent more likely to be engaged in class; in buildings with high-quality professional development, they were 157 percent more likely to be engaged.

Leadership is significantly, but less strongly, related to levels of student engagement in the building. Students in buildings with high-quality SIF, principal and SLC leadership were between 12 percent and 67 percent more likely to be highly engaged. Conversely, the effect of leadership on the likelihood of having low engaged students is mixed. Students in buildings with high-quality principal and SLC leadership were 26 percent and 20 percent less likely to have low levels of engagement, respectively.

Once again, we find that students in buildings with high-quality SIF leadership were 34 percent more likely to experience low levels of engagement. A similar pattern is found for schools with participative decision-making – they are 33 percent less likely to have students who were highly engaged.

Student Perceptions of Teacher Support

Figure XII-8 shows the relationships among the building factors and students' perceptions of how supportive their teachers were. Again, high-quality relationships were most strongly related to students' perceptions of supportive teacher relationships. Students in buildings with high-quality staff and staff-student relationships were 75 percent and 94 percent more likely to report high-quality relationships with their teachers. In buildings with high-quality principal and SLC leadership, students were 15 percent and 23 percent less likely to experience negative teacher relationships. Similar to students' perceptions of implementation and engagement, SIF leadership was negatively related to perceptions of teacher support.

Finally, buildings with high-quality professional development and participative decision-making were more likely to have students who reported supportive teachers and less likely to have students with unsupportive teachers.

Quality of relationships (both among staff, and staff and students) and high-quality professional development are consistently and relatively strongly related to student engagement and support, and to students' perceptions of implementation. SLC and principal leadership was also significantly related to student measures; although the relationship was somewhat weaker and more consistent in reducing negative outcomes. High-quality professional development was also associated with more positive perceptions of instruction and academic/behavioral standards, and with higher levels of engagement and more positive student-teacher relationships. An anomaly was the ratings for SIF leadership, which appeared to be significantly negatively related to student perceptions of instruction, consistent academic and behavioral standards, and engagement and teacher support. One explanation for the link between strong SIF leadership and poor student outcomes and experiences is that the district sends its strongest SIFs to schools that are having more difficulty implementing FTF and engaging their students. This explanation should be pursued in future research.

Collective Engagement and Colleague Support

Figure XII-9 shows the relationships among the building factors and teachers' perceptions of how involved and supportive their colleagues are. Schools with high ratings for six of the seven

Figure XII-6: **Changes in the Percentage of Classrooms Exhibiting Small and Large Grouping Strategies (Classroom Observation)**

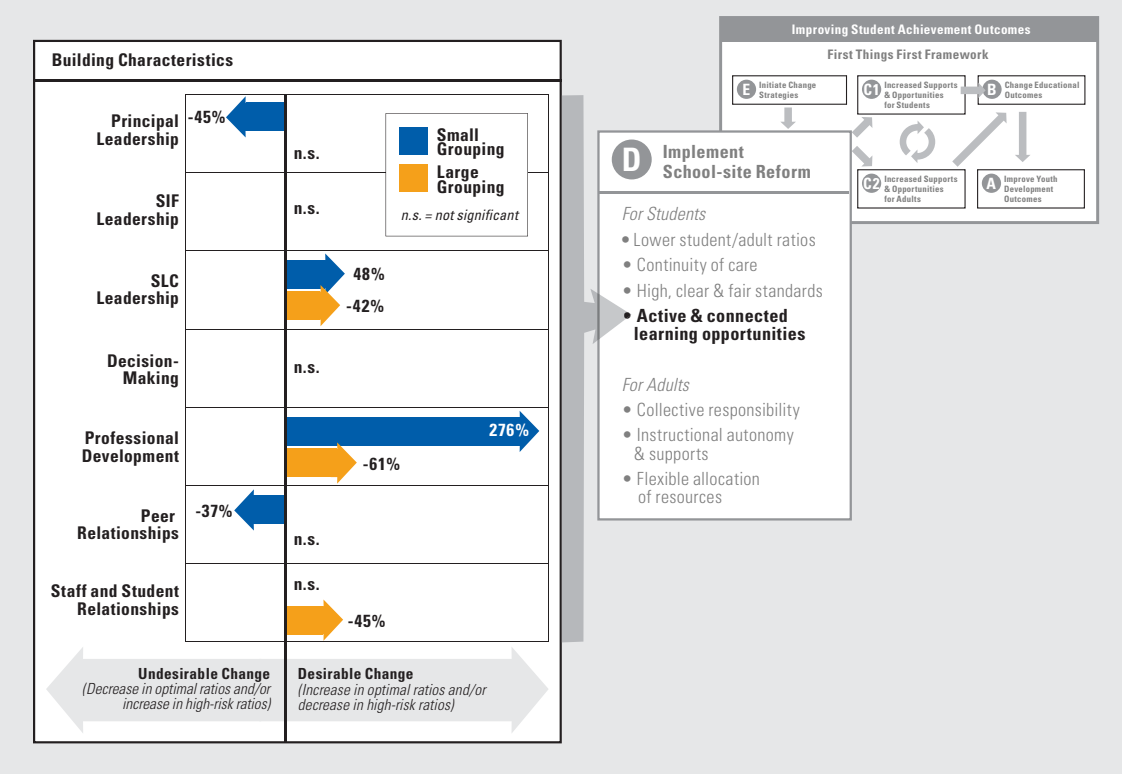
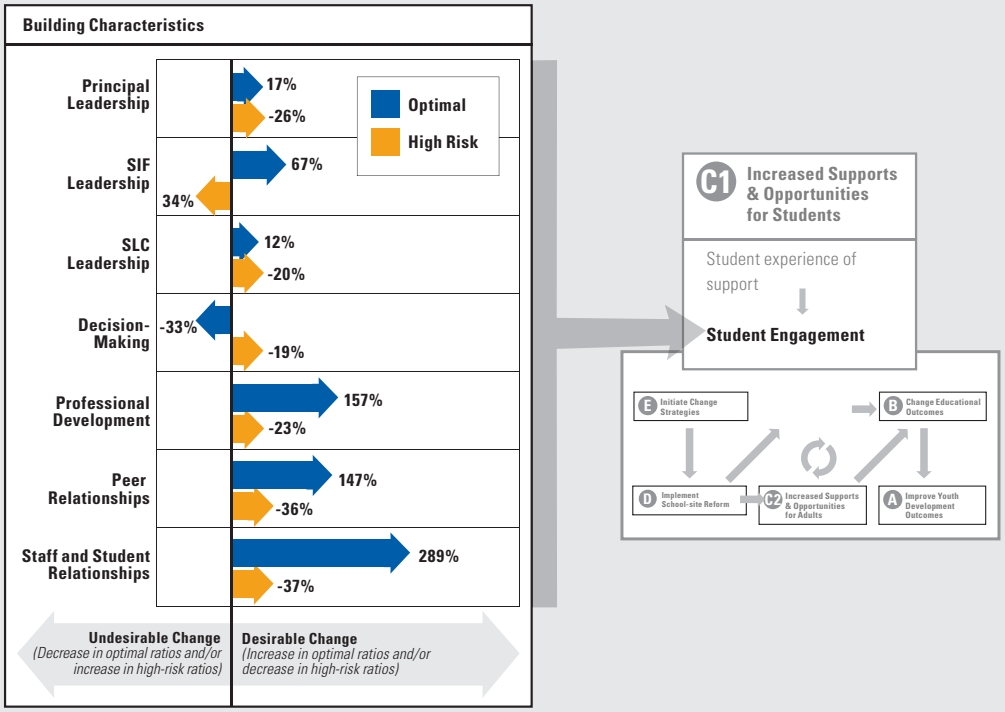


Figure XII-7: **Changes in the Percentage of Students Experiencing Optimal or High-Risk Levels of Engagement (Student Survey)**



implementation factors (principal, SIF and SLC leadership; decision-making; peer support; and staff-student relationships) have significantly more staff who were highly engaged in their work and who reported that their colleagues were highly supportive. Peer and staff-student relationships were most strongly related, with staff members 183 percent and 343 percent more likely to report high levels of collective engagement and colleague support, and 63 percent less likely to perceive their colleagues as disengaged and unsupportive. In addition, schools with strong principal leadership were almost twice as likely to have staff who reported high levels of collective engagement and support, and two times less likely to have staff who reported low levels of collective engagement and support from colleagues. Buildings with high-quality professional development and participative decision-making were also less likely to have teachers who feel that their colleagues were disengaged or unsupportive. Staff members in schools with participative decision-making processes were 64 percent more likely to report that their colleagues were engaged and supportive.

System Support

Figure XII-10 shows the relationships among the building factors and teachers' perceptions of how supportive the administration, school climate and central office are. Significantly more teachers reported high levels of system support in schools rated high in principal and SLC leadership, peer relationships, and staff-student relationships. Peer and staff-student relationships are again most strongly related, with staff members three and four times more likely to report high levels of system support and two times less likely to perceive the administration and school climate as unsupportive. Buildings rated as having high principal and SLC leadership quality were two-thirds more likely to have staff who felt highly supported by the district and building administration, and between one-third and two-thirds less likely to have staff who felt unsupported by the system. Finally, staff in schools with participative decision-making processes were twice as likely to experience high levels of system support, and 40 percent less likely to experience the system as non-supportive.

Staff Engagement

Figure XII-11 shows the relationships among the building factors and teachers' reports of their own individual levels of engagement in teaching. High-quality SLC leadership, participative decision-making and positive staff-student relationships were all associated with a small increased likelihood of staff reporting high levels of engagement (ranging from 23% to 36% increased likelihood for high engagement, and 19% to 31% decrease in likelihood of low levels of staff engagement). In addition, staff in buildings with high-quality SIF leadership and peer relationships were 31 percent and 36 percent more likely to report high levels of engagement, respectively.

It appears that the factors hypothesized as being related to the ability of the schools to implement the reform were significantly associated with staff ratings of colleague engagement and support, and support from the system. Associations between these factors and staff reports of their own engagement were weaker, but were still significant, especially for staff who reported their own engagement to be high. There appeared to be stronger associations between high-quality ratings on the building factors and optimal levels of staff support and engagement than on decreasing high-risk levels of these variables.

Overall, these results suggest that several characteristics of buildings are related to successful implementation of FTF. These quantitative results provide empirical support for the qualitative findings. Staff and student support and engagement, as well as their perceptions of implementation, measured through annual surveys, appear to be most consistently related across all the implementation factors. Some more specific observations and explanations of these findings include:

- The strongest and most consistent results were found for staff variables. It is possible that the implementation factors most directly affected staff within the building, less strongly affected the student variables, and only indirectly affected the building variables. For example, colleague engagement and support, system support, and staff engagement would logically be affected by leadership and relationship factors as we defined them. This interpretation is consistent with the FTF Theory of Change, which posits that staff changes lead to student intermediate outcomes.

- In a number of cases, the SIF leadership ratings did not behave as expected. Specifically, lower student ratings of teacher support and engagement tended to be associated with higher ratings for SIF leadership; higher student ratings of teacher

Figure XII-8: **Changes in the Percentage of Students Experiencing Optimal or High-Risk Levels of Teacher Support (Student Survey)**

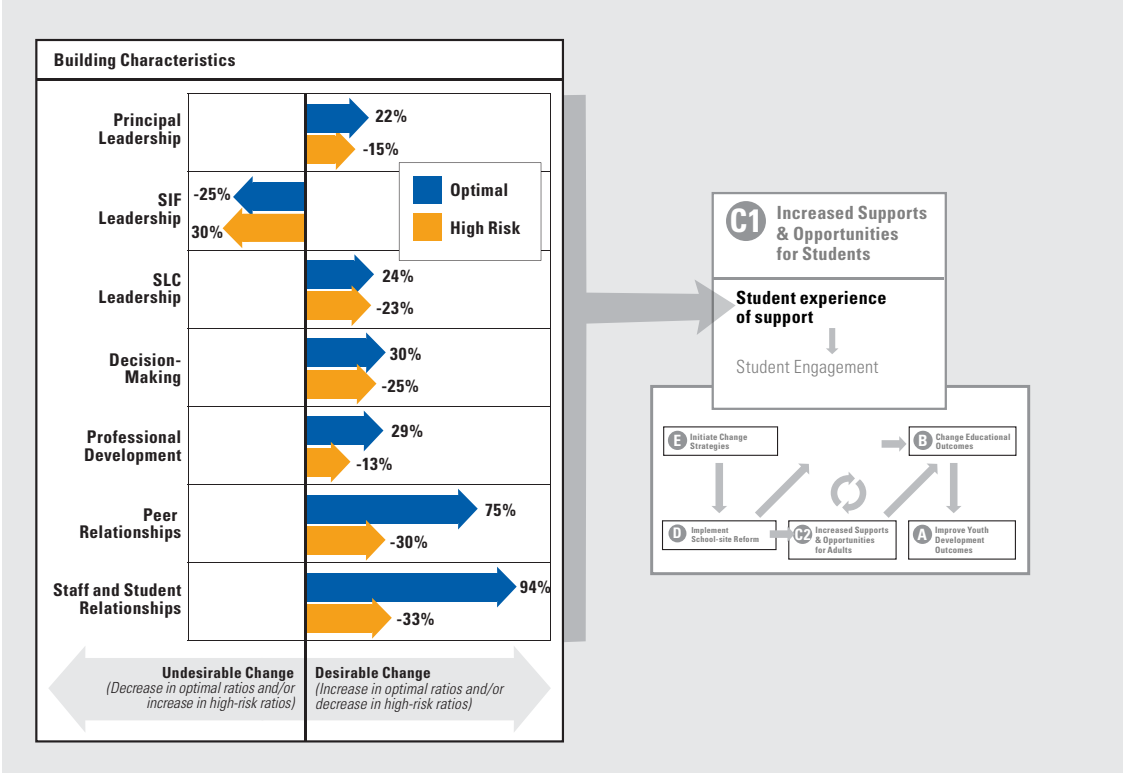


Figure XII-9: **Changes in the Percentage of Staff Experiencing Optimal and High-Risk Levels of Collective Engagement and Colleague Support (Staff Survey)**

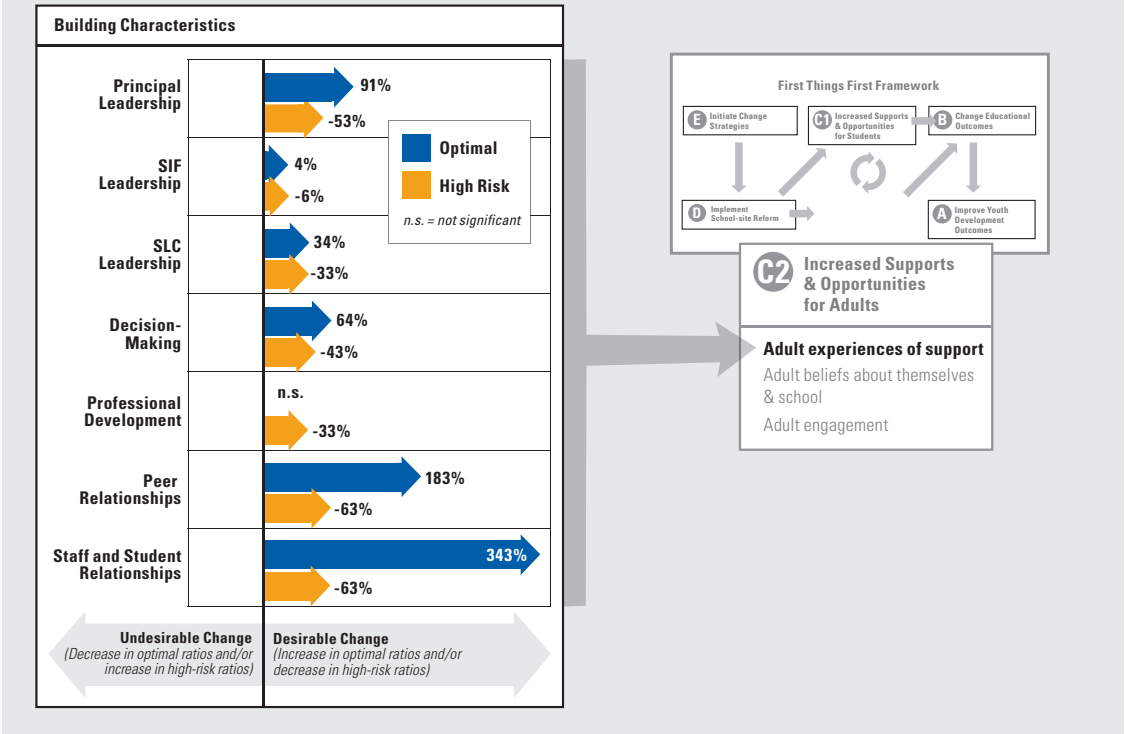


Figure XII-10: **Changes in the Percentage of Staff Experiencing Optimal and High-Risk Levels of System Support (Staff Survey)**

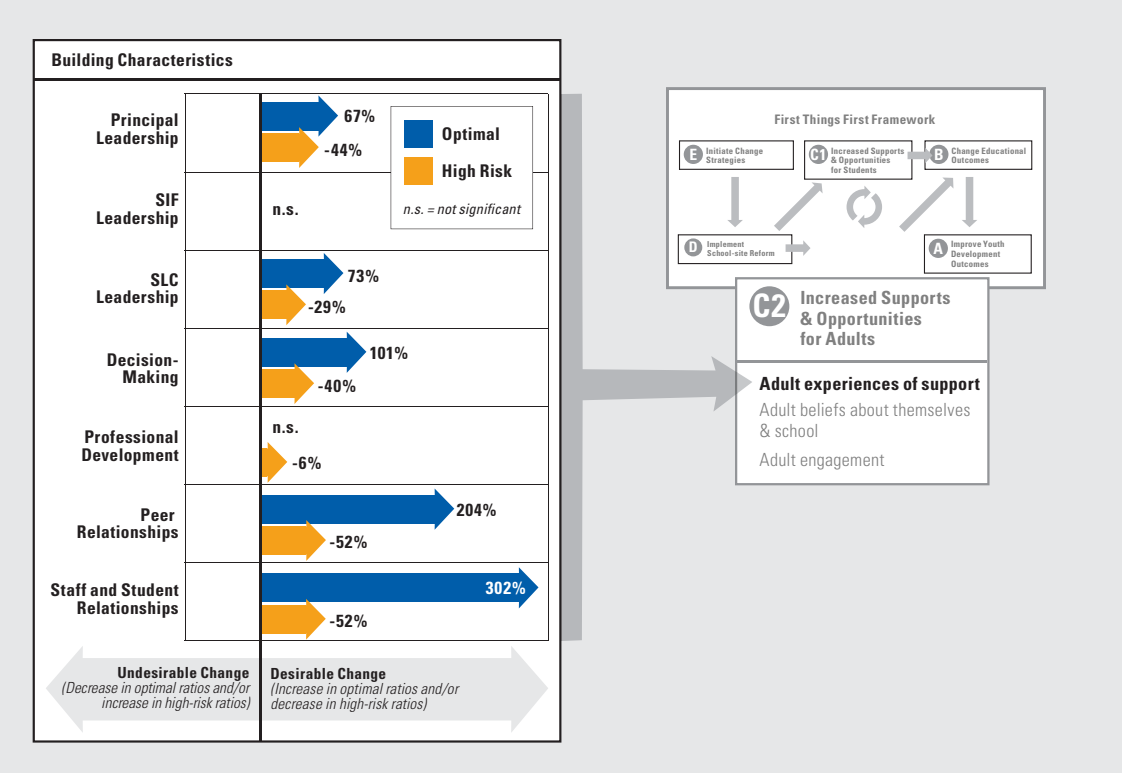
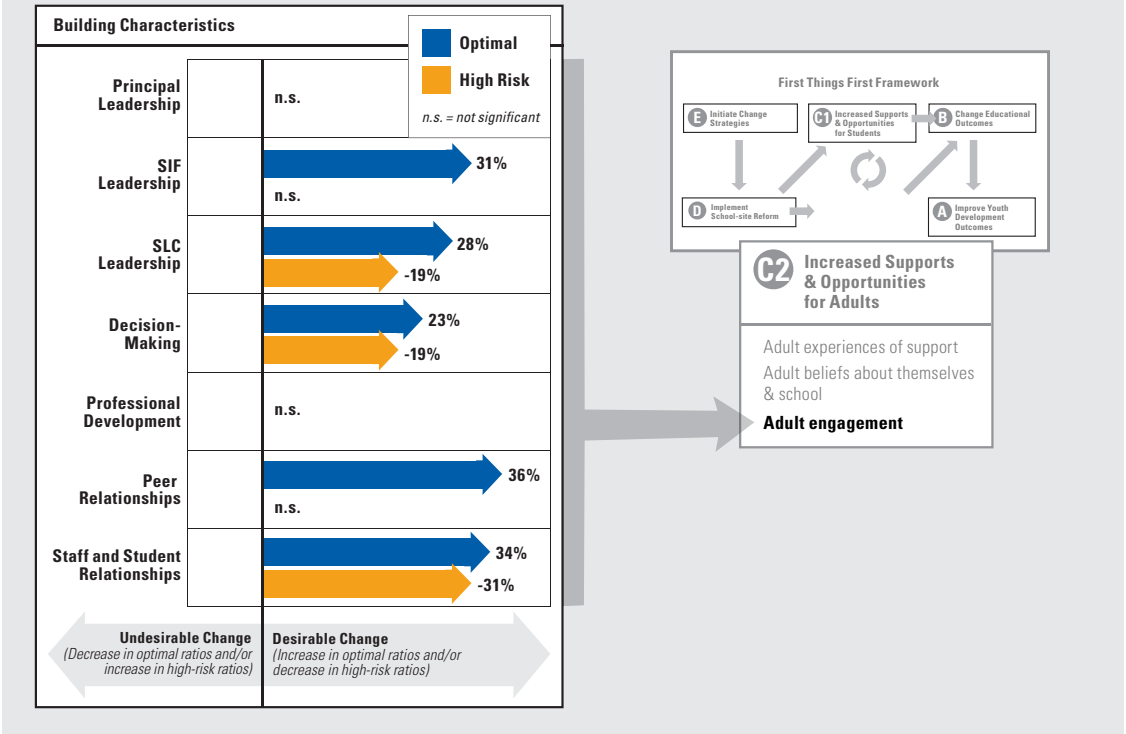


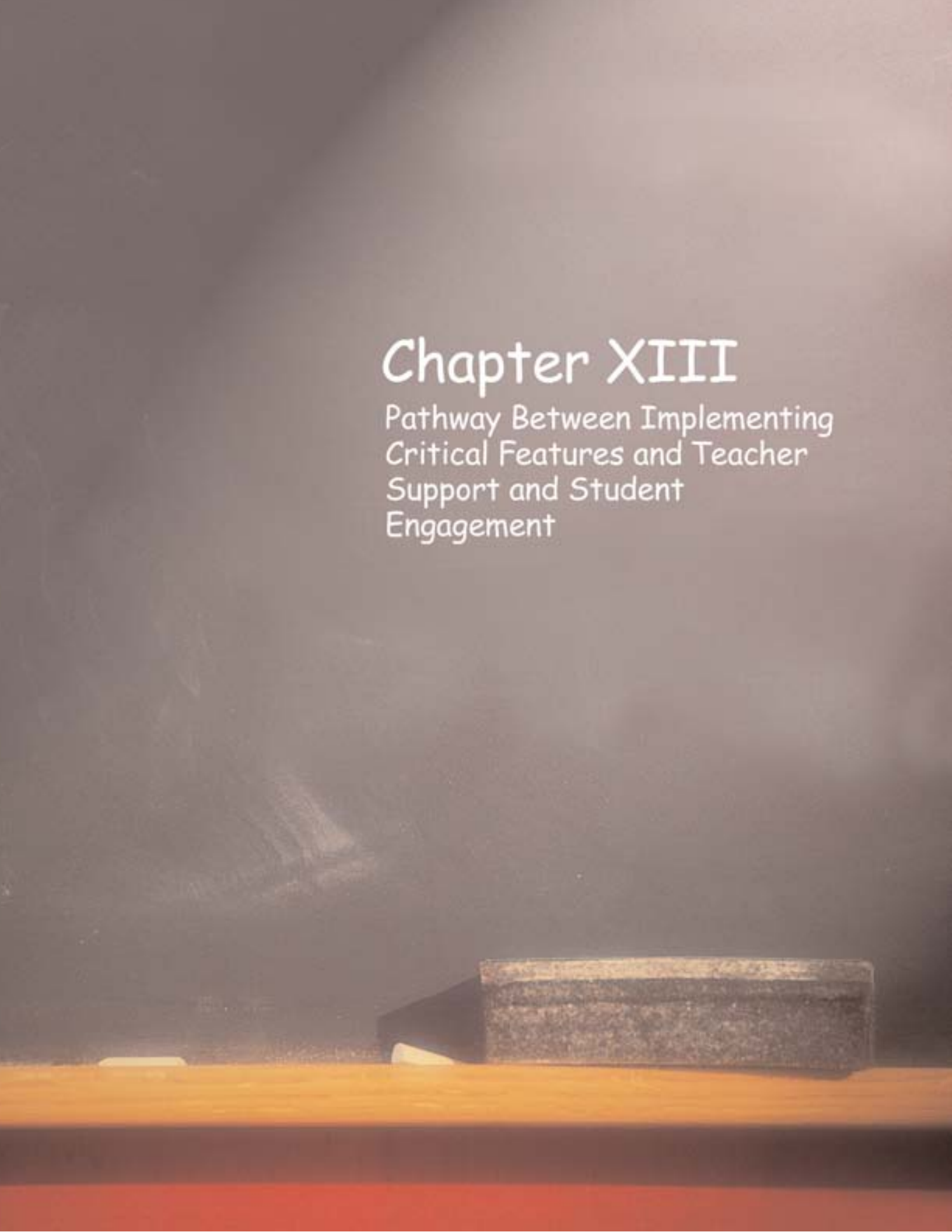
Figure XII-11: **Changes in the Percentage of Staff Experiencing Optimal and High-Risk Levels of Staff Engagement (Staff Survey)**



support were associated with lower ratings for SIF leadership; and failure to observe lower student-adult ratios were associated with higher ratings for SIF leadership. Finally, in the case of SLC stability and teacher ratings of systems support, we found significant relationships between high SLC stability and both low and high ratings of SIF leadership. These findings may reflect management decisions rather than an actual influence of the SIF on outcomes. That is, SIFs with greater leadership skills may have been assigned to schools that were seen by the management team as struggling.

- As noted, some of the results appear to reflect a "dip" in these observation-based variables that occurred in the 2001-2002 data collection. During that year, several converging events may have resulted in reduced attention to the Wyandotte and Washington clusters: (1) in the 2001-2002 year, two newly hired executive directors were assigned to these clusters; (2) several of the most experienced SIFs were reassigned as instructional coaches, with responsibilities across seven to eight schools rather than the two schools that were their previous assignment, and a single school that was their assignment in 2002-2003; and (3) IRRE was involved at a much lower level during the 2001-2002 year, with most of their resources focused on the central office rather than in the buildings.

Most of the factors derived from the qualitative study of KCK schools as they engaged in planning and implementing FTF, are empirically related to the intermediate outcomes of this evaluation. This finding holds promise for future replication of the reform model. The specific indicators for each factor appear to form a relatively objective basis for judgment as to whether such intangible qualities as "principal leadership" or "peer relationships" are in place. If these factors can be reliably measured, this means they may also be reliably taught in professional development initiatives. Further, administrators may be able to base their management decisions and personnel assignments on some objective criteria designed to enhance the indicators related to the factors we believe lie behind successful implementation of this reform initiative.



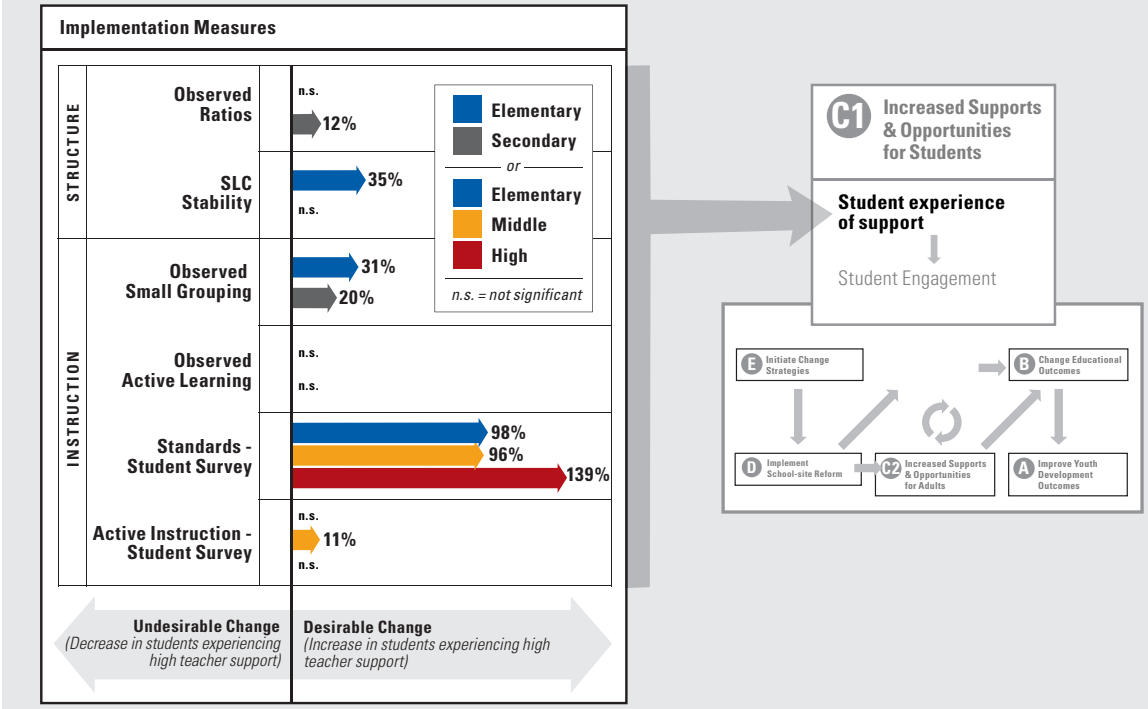
Chapter XIII

Pathway Between Implementing Critical Features and Teacher Support and Student Engagement

MAJOR FINDINGS

- *There was clear evidence for the link between implementing high-quality structure and instruction and students' experience of support and engagement.*
- *The strongest link in the model was between students' perception of standards and how supported they felt by their teachers. Elementary and secondary students who experienced high academic and behavioral standards were twice as likely to feel supported by their teachers as the average student. Secondary students were also more likely to be engaged if they experienced consistently high academic and behavioral standards.*
- *In the two clusters implementing FTF for more than three years, elementary students who were in a SLC for two or more years were more likely to feel close to their teachers; secondary students were more likely to report being engaged in school.*
- *Secondary students in the Wyandotte and Washington clusters were also more likely to report having highly supportive teachers and less likely to be disengaged when learning in lower student-adult ratios.*
- *Wyandotte and Washington students at all levels, who were instructed in smaller grouping formats, were more likely to feel supported by their teachers*
- *There was little connection between observed active learning or student reports of active instruction and intermediate outcomes. This may be due to the inconsistent implementation of active instruction over time and across classrooms, or it could be due to the inability of the measurement tools to fully capture the active learning experience.*

Figure XIII-1: **Increase in Percentage of Students Experiencing High Teacher Support**



PATHWAY BETWEEN IMPLEMENTATION OF FIRST THINGS FIRST AND INTERMEDIATE OUTCOMES

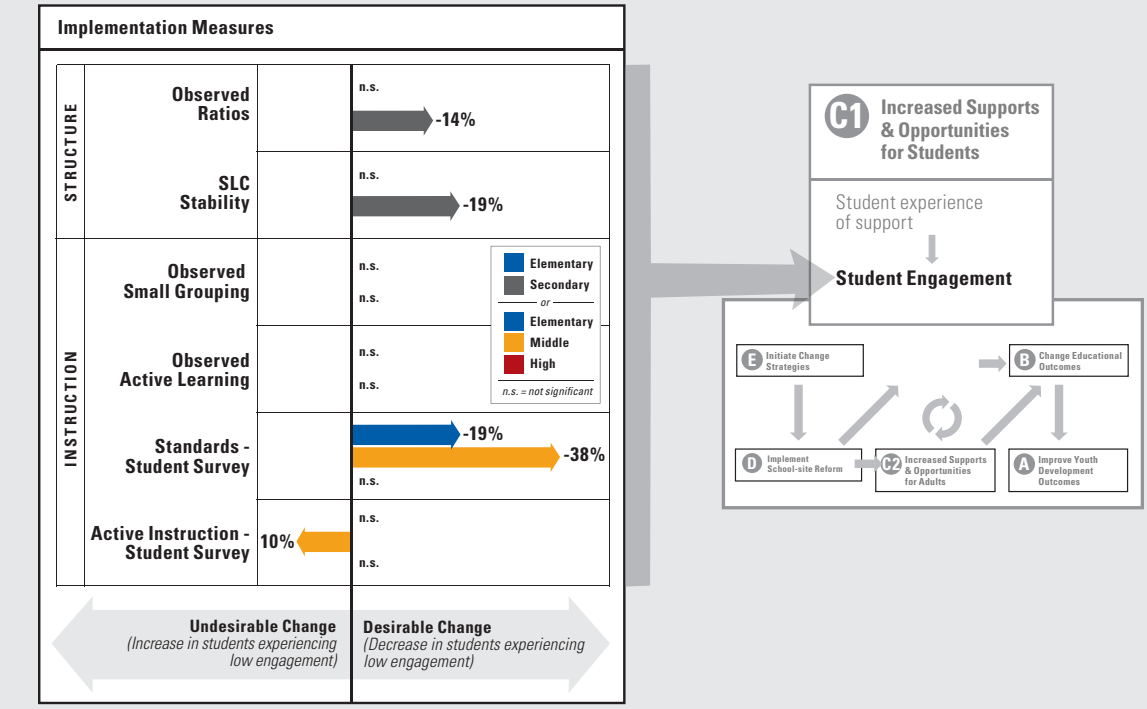
This chapter examines the relationships between implementation and intermediate outcomes. Findings from across the separate analyses that were conducted for each outcome are summarized here. This allows us to compare the relative strength of pathways between different sets of outcomes. These models are explained in greater detail in Appendix A.[26]

RELATIONSHIP BETWEEN IMPLEMENTATION AND TEACHER SUPPORT

Figure XIII-1 presents the relationship between each of the implementation measures and students' experience of high teacher support. Each figure presents the relationship between students' experience of one outcome (e.g., ratios) and their experience of a second outcome (e.g., teacher support). The numbers in the figures indicate the increase in the likelihood of students experiencing optimal levels of an outcome when they also experienced optimal levels of the first outcome. For example, secondary students

[26] When the dependent variable comes from observation data, the results are split by elementary and secondary levels; when it comes from survey data, we split the results by elementary, middle and high school levels. The difference in how levels are defined is because the observation data were collected at the classroom level, and the sample size does not allow us to estimate separate models for middle and high school.

Figure XIII-2: **Decrease in Percentage of Students Experiencing Low Engagement**



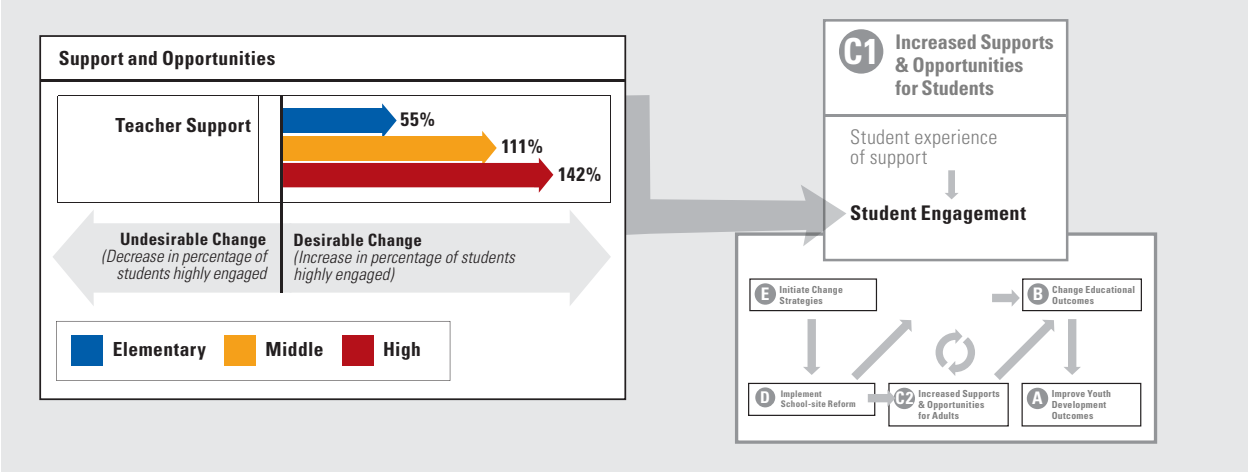
were 12 percent more likely to experience positive relationships with teachers if they learned in optimal student-adult ratios.

The relationships between the measured implementation variables on students' experience of high levels of teacher support were relatively small, with the exception of standards. Students who experienced consistently high academic and behavioral standards were between 96 percent and 139 percent more likely to experience positive, supportive relationships with their teachers. Effects of structure and instruction were smaller. Elementary students were 35 percent more likely to have positive teacher relationships when they were involved in SLCs for two or more years. Students who experienced high levels of small grouping instruction in the classroom were between 20 percent and 31 percent more likely to have positive relationships with their teachers, and secondary students were 11 percent more likely to have positive relationships with their teachers when they experienced high-quality instruction.

RELATIONSHIP BETWEEN IMPLEMENTATION AND ENGAGEMENT

As shown in Figure XIII-2, the relationships between the implementation measures and low levels of engagement were relatively small. Once again, the experience of high academic and behavioral standards has the strongest association with the likelihood of a student being disengaged from school – this is particularly evident for high school students who were more than one-third less likely to be disengaged if they had high standards. In addition, secondary students who experienced lower teacher-student ratios were 14 percent less likely to be disengaged; they were 19 percent less likely to be disengaged when they were in a SLC for two or more years. Observed instructional practices did not appear to be related to high levels of student engagement; however, middle school students who reported experiences of active instruction were 10 percent more likely to be disengaged. One possible explanation for this finding is that the active

Figure XIII-3: **Increase in the Percentage of Students Highly Engaged**



instruction strategies, such as cooperative and project-based learning, were not in place for long enough to re-engage students who were chronically disengaged (e.g., at-risk youth).

The results clearly indicate that the implementation variables are more strongly linked to teacher support than to engagement. This finding is consistent with the FTF theory of change framework which posits that implementation has an indirect effect on engagement through teacher support. This linkage from classroom experiences to teacher support to engagement has been supported in the literature as well (e.g., see National Research Council & Institute of Medicine, 2004, Chapter 2).

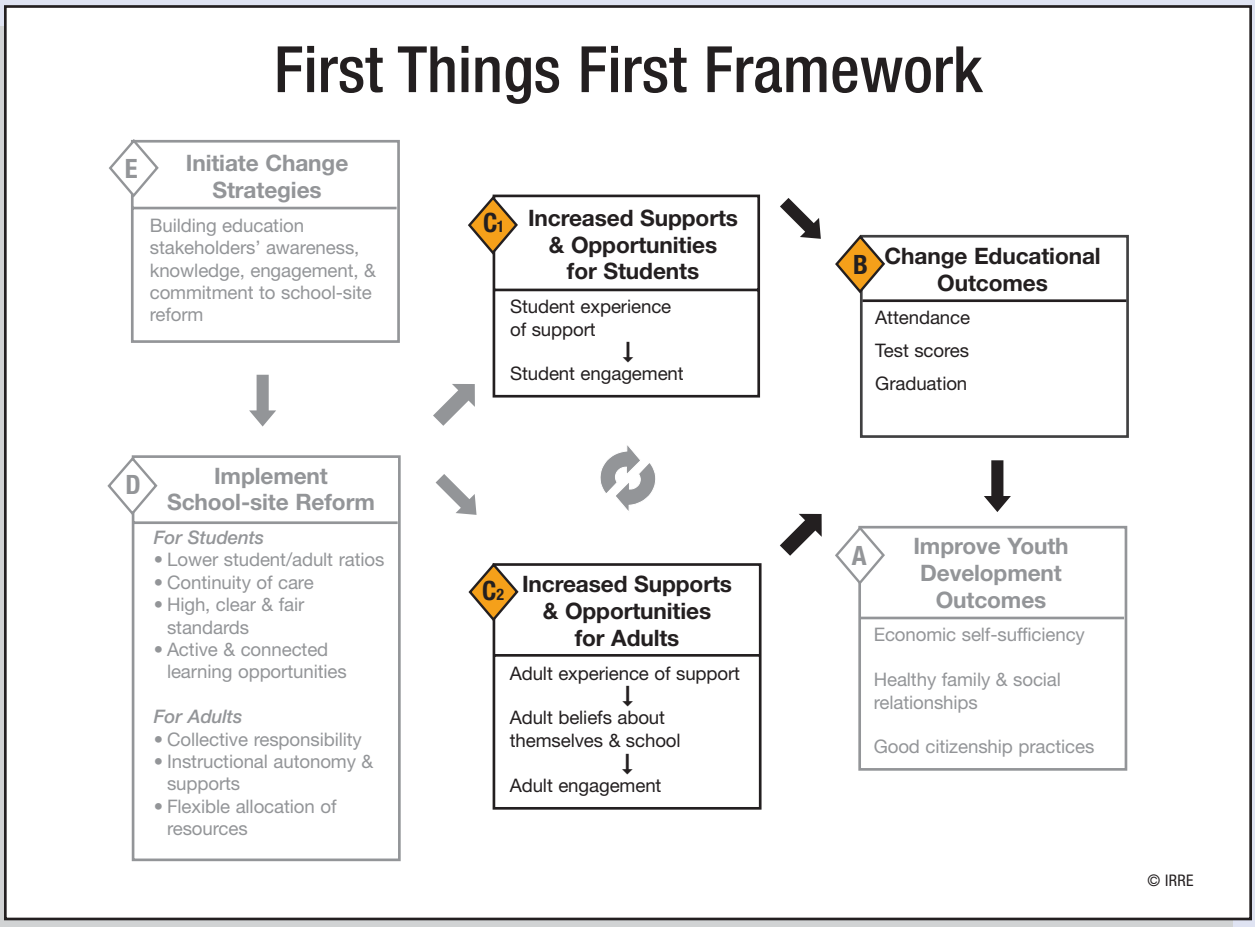
RELATIONSHIPS AMONG INTERMEDIATE OUTCOMES

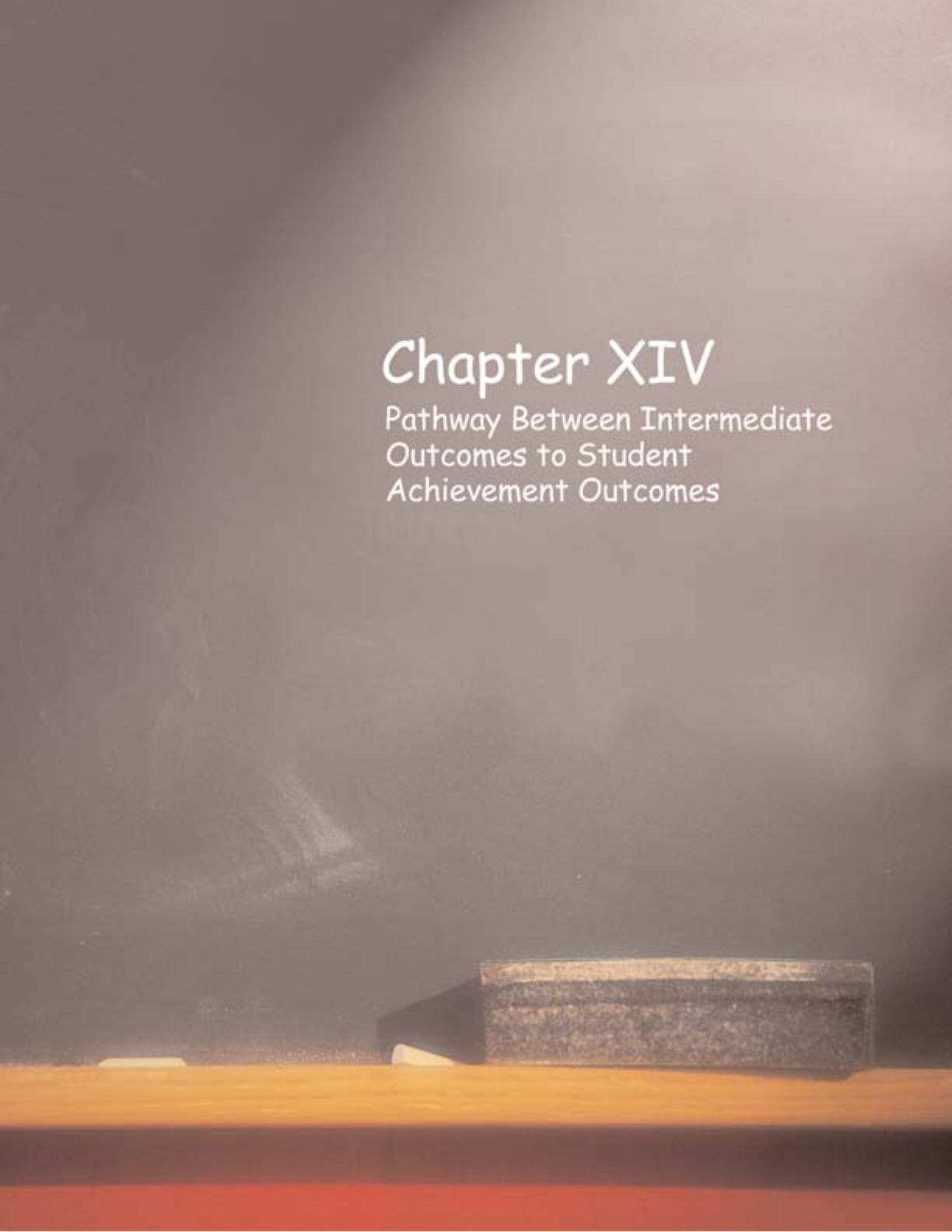
Figure XIII-3 shows the relationship between teacher support and student engagement in the classroom. These findings validate the relationship between teacher support and student engagement previously established in the literature (e.g., Willingham, Pollack & Lewis, 2002; Ryan & Patrick, 2001; Marks, 2000; Solomon, Battistich, Watson, Schaps & Lewis, 2000; Finn & Rock, 1997; Connell, Halpern-Felsher, Clifford, Crichlow & Usinger, 1995; Voelkl, 1995). However, because we are able to indicate how many more students would be engaged in school if they experienced supportive teachers, we believe these results are more useful and compelling for school stakeholders and policymakers.

Students that had teachers who were fair, cared about how they did in class, and listened to and liked them, were twice as likely at the middle school level and two and a half times as likely at the high school level to be engaged in school. Although elementary students who had supportive teachers also were more likely to be engaged, the relationship was not nearly as strong. This variation could be due to secondary students being less likely to experience supportive environments than elementary students.

There were virtually no significant relationships between staff experiences of their own support and engagement and student engagement. It is interesting to note, however, that in schools where colleagues were supportive of one another, students were less likely to have poor relationships with their teachers.

QUESTION 6: Are improved intermediate outcomes (supportive relationships and engagement in school) associated with better long-term outcomes for students?





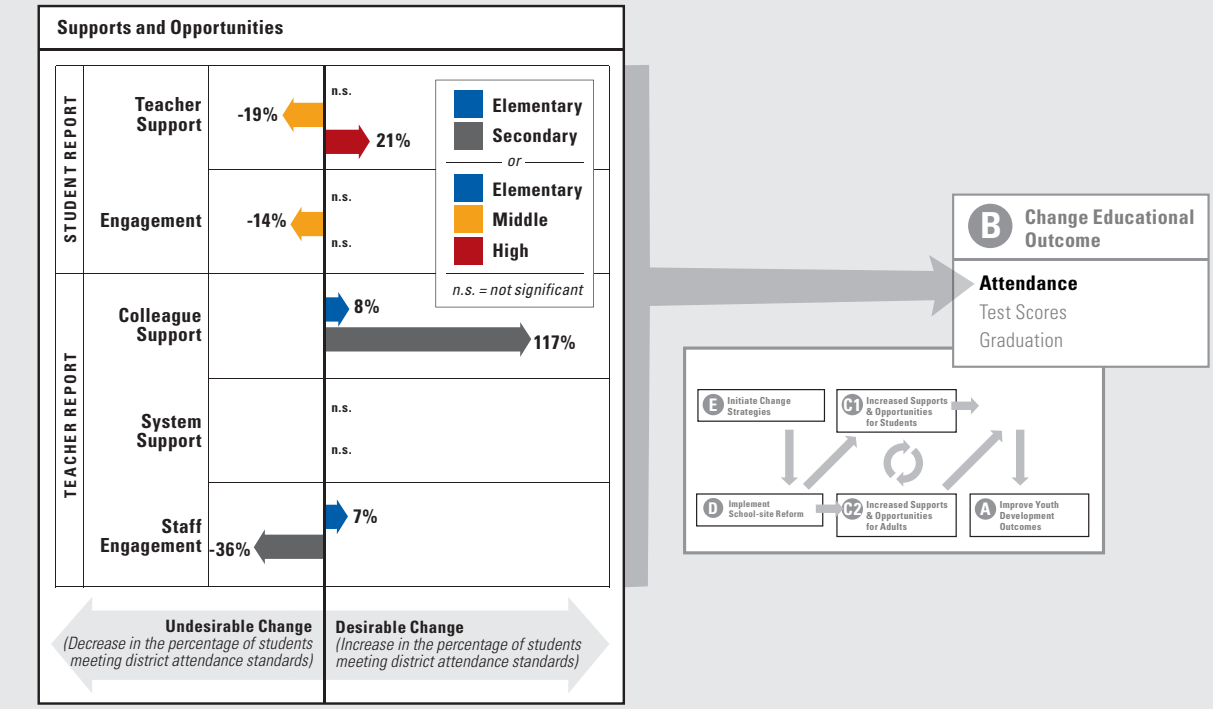
Chapter XIV

Pathway Between Intermediate Outcomes to Student Achievement Outcomes

MAJOR FINDINGS

- *Students who felt supported by their teachers were much more likely to score proficient or above on the math and reading portions of the state test. Teacher support also had a strong positive association with the likelihood that secondary students would meet the district standard for attendance.*
- *Middle school students who were engaged were more likely to meet district attendance standards. Elementary students who were engaged in school were more likely to score proficient or higher on the state reading and math tests.*
- *Students were more likely to have high levels of attendance in schools where staff members felt supported by their colleagues.*
- *Elementary schools in which staff members reported being engaged in their teaching were more likely to have students with good attendance and high state test scores in reading. Secondary schools with highly engaged staff were more likely to have students score proficient or above on state math tests.*

Figure XIV-1: **Increase in the Percentage of Students Meeting District Attendance Standards**

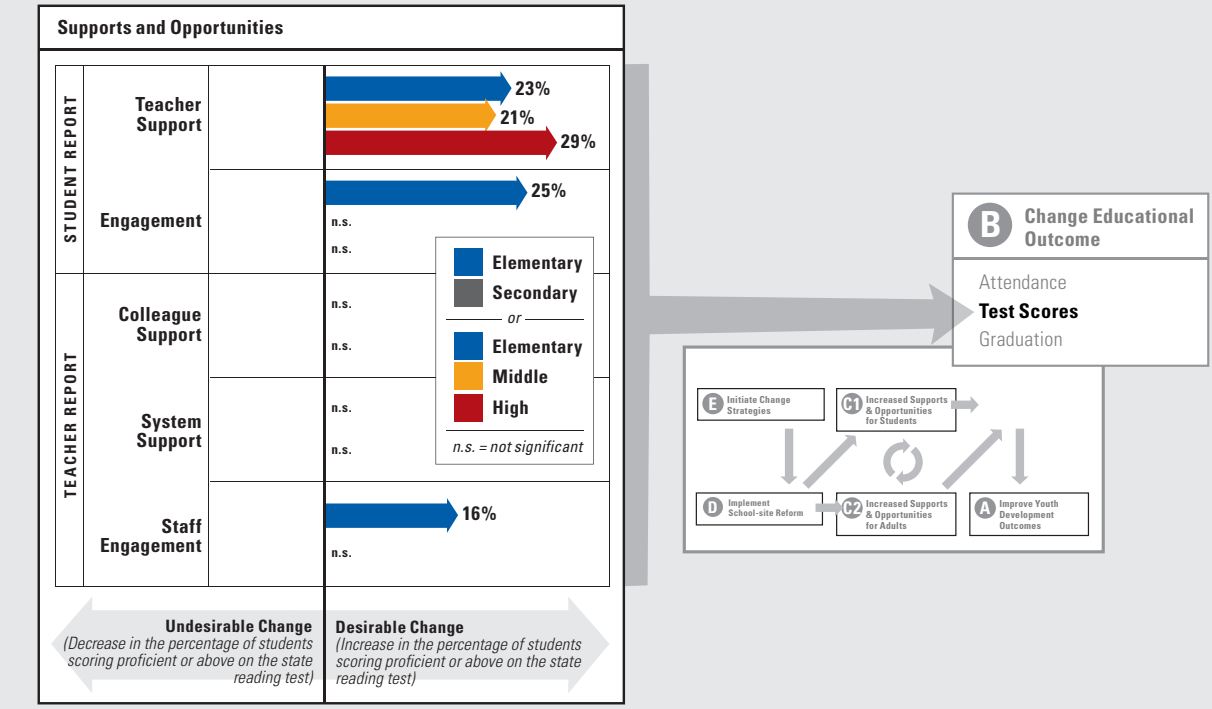


RELATIONSHIP BETWEEN INTERMEDIATE OUTCOMES AND ATTENDANCE

The relationship between student and staff experiences of support and engagement and student attendance is presented in Figure XIV-1. There were few relationships between the intermediate outcomes and elementary student attendance because, similar to most elementary students, KCK grade school students had consistently high levels of attendance. Thus, there is little room for improvement in the attendance rates of elementary students.

In contrast, secondary students' attendance did significantly improve when they experienced supportive teachers, felt more engaged and attended schools that have supportive staff members. It is particularly interesting to note that students in secondary schools in which staff members reported high levels of support from their colleagues (e.g., shared resources, helped and encouraged each other) were twice as likely to meet the district attendance standard. It is possible that teachers who have collegial relationships are more likely to discuss individual students' problems (e.g., absenteeism) and strategies for resolving those problems. It is unclear why there was a negative relationship between staff engagement and student attendance at the secondary level.

Figure XIV-2: **Increase in Percentage of Students Scoring Proficient or Above on the State Reading Test**



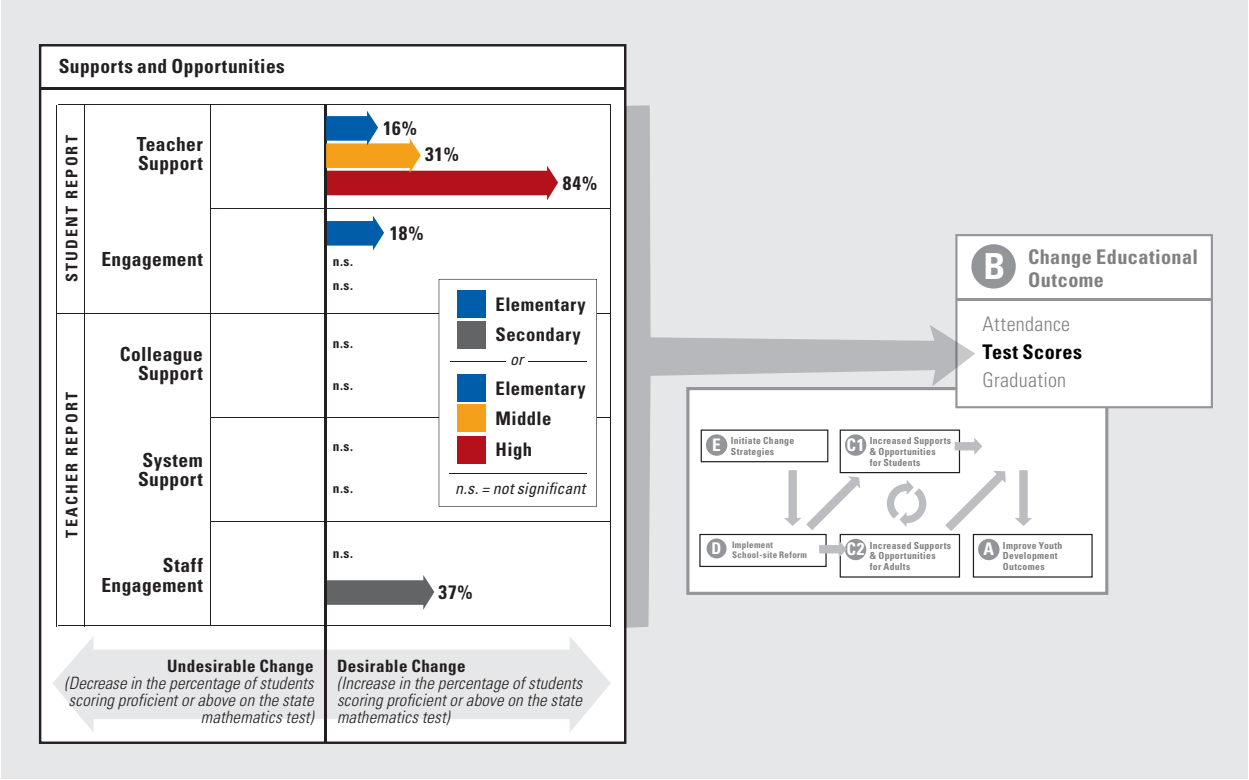
RELATIONSHIP BETWEEN INTERMEDIATE OUTCOMES AND STATE TEST SCORES

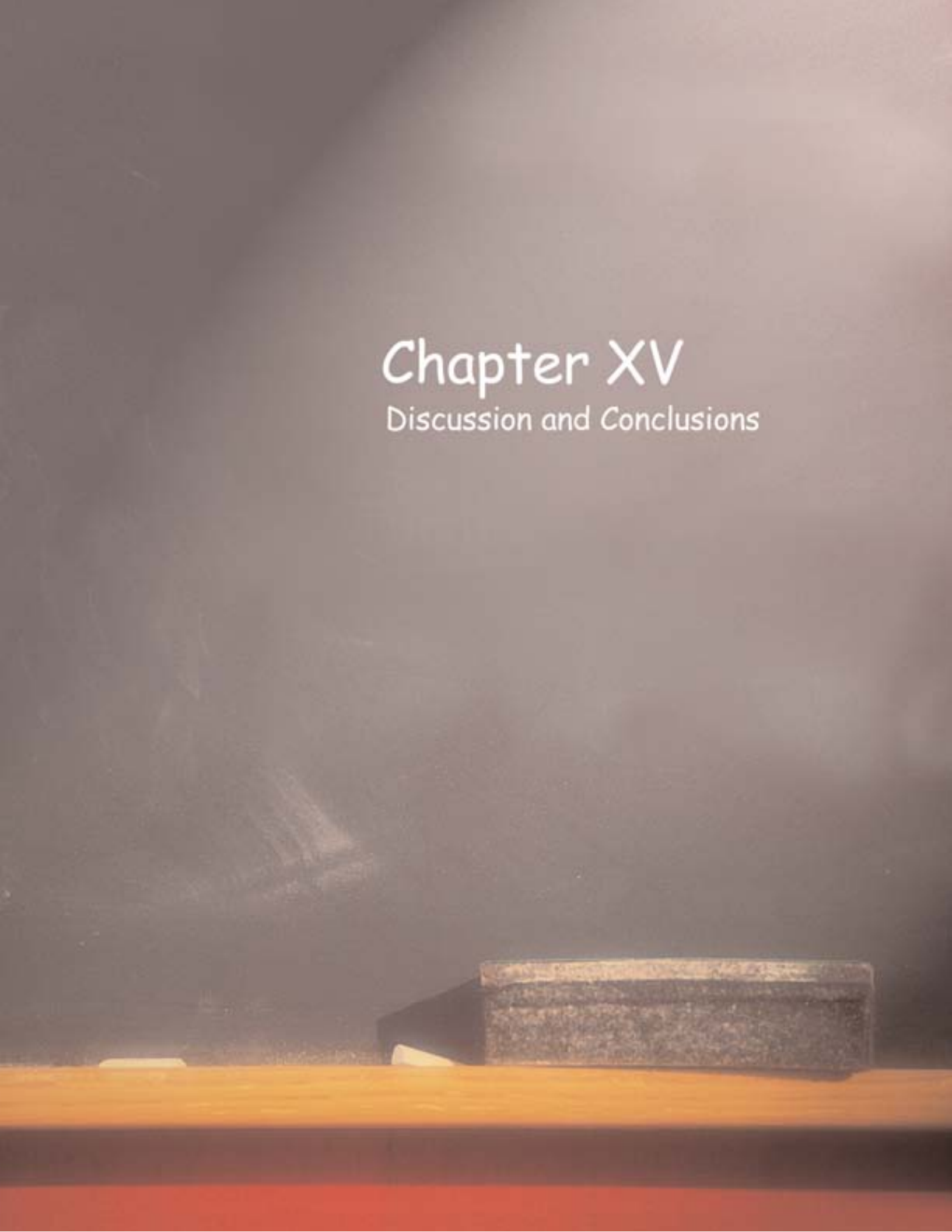
As shown in Figures XIV-2 and XIV-3, the experience of supportive teachers continues to have the strongest link to student outcomes. Students who reported having strong relationships with their teachers were approximately 25 percent more likely to score proficient or above on both the state reading and math tests. This holds at all levels – elementary, middle and high.

In addition, elementary students were more likely to be proficient or better on state tests if they reported being engaged in school. Finally, when staff were engaged in their work, students were also more likely to score proficient or above on state tests (with 16% more elementary students scoring at or above proficient on the reading portion, and 37% more secondary students scoring at or above proficient on the math portion).

The results clearly provide support for the linkages in the FTF framework. Measures of implementation influenced intermediate outcomes which, in turn, influenced long-term outcomes. Measures of implementation had the strongest association with teacher support, which, in turn, demonstrated clear relationships with engagement, attendance and test scores. Student engagement also seems to be a mediating variable between implementation and long-term outcomes, but for fewer outcomes than teacher support.

Figure XIV-3: **Increase in Percentage of Students Scoring Proficient or Above on the State Mathematics Test**





Chapter XV

Discussion and Conclusions

Sustaining a comprehensive, district-wide reform effort in an urban school district is relatively rare; the evaluation associated with this effort is also relatively uncommon in its scope and longevity. As such, it holds important lessons for policymakers, practice stakeholders and researchers.

The Ewing Marion Kauffman Foundation made an early and substantial commitment to evaluation of the reform, allowing us to plan for rich, theory-driven, longitudinal data collection that measured both the progress and effectiveness of the initiative. We were able to document changes at the district and building levels, have multiple measures of implementation, and track progress on improving both intermediate and long-term outcomes over a period of six years. As a result, we were able to test each of the hypothesized steps in the FTF theory about what it takes to improve conditions in all schools in a district.

JUDGING THE EFFECT OF FIRST THINGS FIRST

Over time, some shortcomings in the data available arose due to the inaccuracy of district records prior to the reform (e.g., attendance data) and changes in the data collection design required by political realities within the district (e.g., staff surveys, test scores) that we addressed in our analyses. As a rule, wherever a shortcoming in our data existed, we chose a strategy for addressing it that provided the most conservative estimate of change or association.

The one design element that would have strengthened our ability to estimate the effects of FTF on outcomes is a control/comparison group. It is possible to randomly assign school buildings to a reform (e.g., Cook, Murphy & Hunt, 2000; Kemple & Snipes, 2000), but it may never be realistic to randomly assign whole school districts to treatment or control groups. However, even with district-wide reform, ideally it is possible, in some cases, to select carefully matched comparisons. In Kansas City, Kansas (KCK), the practical realities of adapting to unfolding events prevented us from implementing such a design. We had planned to use the later implementing clusters as a comparison group for the earlier clusters, but when it became clear that those clusters were starting to make changes on their own and the implementation plan was accelerated,[27] it was apparent this component of the design would not work.

Fortunately, the embeddedness of the research in the initiative and the addition of research funding by the U.S. Department of Education (DOE)[28] allowed us to extend the period of the evaluation long enough to collect data through the third year of implementation for the last two implementing clusters. The lengthier data collection period and staggered implementation enabled us to assess change associated with each additional year of implementation of FTF for three separate groups of schools, over three years of implementation, occurring in three different historical timeframes. We used this change in implementation to develop one of our "control" strategies when analyzing the data.

In addition to using analytic techniques to minimize the likelihood that findings could be due to factors extraneous to the reform effort (i.e., demographic changes in the student population or historical contextual changes in the district, state or related policies), we also need to consider the overall

[27] This phenomenon is described and discussed in the previous evaluation report on this initiative – see Gambone, Klem, Moore & Summers, 2002.

[28] When the DOE chose FTF as one of seven models to undergo further development and testing, additional resources were available for the KCKPS initiative.

pattern of findings in assessing the contribution of FTF to the improvements seen in the district. Especially in circumstances where randomized control group studies are not feasible, one important and powerful strategy for testing an initiative's effectiveness is to examine whether the changes predicted by the theory of change occur in the expected manner (Granger, 1998; Weiss, 1995).

THE PATTERN TO DATE

The overall pattern of findings presented here showed that:

The Critical Features of FTF Were Implemented

- Each additional year of FTF implementation was associated with greater proportions of students having the type of educational experience outlined by FTF – smaller student-teacher ratios; continuity in Small Learning Communities (SLCs); high academic and behavioral standards; and, at the secondary level, more opportunities to participate in active learning.
- Several factors were related to stronger implementation – principal and SLC leadership, positive relationships among staff members and between staff and students. Participatory decision-making was related to smaller classes and continuity of SLCs, whereas high-quality professional development was related to instructional changes.

Students in Buildings with More Implementation of FTF's Critical Features had Better Intermediate Outcomes – Support and Engagement

- The more years a building implemented FTF, the greater the proportion of students who received the type of support and caring from teachers that research shows to be so important to achievement – particularly in urban districts. We also saw a drop in the proportion of students highly disengaged from school with longer implementation of FTF.
- High levels of academic and behavioral standards had the strongest relationship with the likelihood of students experiencing support from their teachers. Students experiencing high standards were also less likely to feel disengaged from learning. Learning in small group formats was also associated with better relationships with teachers.
- Secondary students were more likely to feel supported by their teachers and less likely to be disengaged from school when they had lower student-adult ratios and were in SLCs for at least two years.
- Students who felt supported and cared about by their teachers were more engaged in school. This relationship was strongest for secondary school students.

When Intermediate Outcomes Improved, Student Achievement Improved – Students Came to School More, Performed Better, and Were More Likely to Graduate on Time and Less Likely to Drop Out

- Students who felt supported and cared for by their teachers were much more likely to score at or above proficiency levels on the math and reading portions of the state assessment. Secondary students who had good relationships with their teachers also were more likely to attend school on a regular basis.
- With the advent of the No Child Left Behind Act (NCLB), the highest stakes outcome examined in this study was student performance on the state test. Not only did the district as a whole show more improvement on the state test than did the rest of the state, but the gap between minority and non-minority groups also closed faster in KCK than it did in the rest of the state.

- Higher levels of engagement were not as strongly associated with improved long-term outcomes as was teacher support; however, elementary students who reported being engaged in school were more likely to score at or above proficiency levels on the state tests in both reading and math. And middle school students who felt engaged had higher levels of attendance.

- Finally, students were more likely to meet the district attendance thresholds in schools with high levels of collegial support among staff, particularly at the secondary level. In addition, schools with highly engaged staff were more likely to have students score at proficiency levels on the state math assessment (secondary) and on the state reading assessment (elementary).

This pattern of findings closely follows the FTF theory with two exceptions. First, the improvement that occurred by school year 2002-2003 in student-adult ratios came from moving youth out of the "risk," or lowest level, rather than moving more students into the highest level. Second, two improvements were seen only at the secondary level – increased active learning strategies and increased staff perceptions of support from their colleagues. Otherwise, we found every improvement hypothesized by the theory, across both elementary and secondary schools.

This pattern, in conjunction with the statistical controls used in the analyses, leads us to conclude that it is likely that FTF was the vehicle for these improvements. It is possible that simultaneously increasing external pressures on the district to improve (e.g., state testing mandates and NCLB) may have also increased the degree of urgency felt by some in the district to engage in the reform. But the means and methods for the changes were clearly furnished by a strong reform model and the collaboration of the school district, technical assistance partner and funding partner that formed the initiative's leadership.

Given these results, it is also helpful to examine what we have learned about what factors seem most important in producing this kind of change in a district-wide reform effort.

THE ROLE OF A DISTRICT IN REFORM

Instituting change from the top down is a difficult proposition. Previous research on the elements needed for successful comprehensive reform has shown that externally developed models are more effectively implemented, and have better results, than those that rely on internal development. The research also shows, however, that in order for reforms to be successfully sustained, teacher buy-in and a sense of ownership are also critical components for establishing the legitimacy and credibility of a reform (Desimone, 2002; Berends, Chun, Shuyler, Stockley & Briggs, 2002). We explored in detail in an earlier report (Gambone, Klem, Moore & Summers, 2002) the strategies employed by this initiative to address the challenge of providing both the "specificity" and "authority" that others have shown to be the hallmarks of successful system-wide change (Anderson, 2003; Desimone, 2002; Porter, 1994).

The underlying challenges of reform of any system arise from the fact that people have a natural tendency to resist change, and that large systems and organizations are difficult to reform. Three themes, interwoven across the strategies employed by the central office to meet these challenges offer insight into some of the conditions necessary for achieving "top-down" change: *commitment, consistency* and *clarity*.

Commitment

District leadership at all levels showed visible commitment to the FTF initiative. Administrators

demonstrated that they were willing to engage in restructuring processes that were as difficult as those they were asking the schools to undertake. In the same way that teachers were asked to change the way they did business every day, many of the roles of central office staff were changed. Leaders also made their involvement in the reform noticeable to the schools by participating in learning and development opportunities along with staff, and providing tangible support and resources to further the goals of instruction and reform.

Significant resources were also dedicated to developing the capacity of principals to lead the reform in their buildings – increasing their engagement in the reform. The unions were engaged early in the process and backed the changes required by the reform. And, the School Board demonstrated commitment to the reform in key decisions – by explicitly selecting a new superintendent based on his dedication to FTF and by approving an early release policy one day a week so that staff could engage in professional development. This signaled to all staff that all levels of the district's leadership were committed to seeing the reform happen.

Consistency

All administrators were "on message" (not an easy process to achieve) with the goals of FTF. All of the changes made in the district were addressed within the context of the FTF framework until the reform was seen as the fundamental work of the district rather than a short-term change effort. This consistency was important to maintain across the years, thus chipping away at the "this too shall pass" mentality of those administrators and teachers whose usual response to change was to hunker down and wait for the initiative to go away.

Further, at both the district and building levels, FTF provided a mechanism for organizing efforts; and, in a sense, protected staff from being buffeted by emerging outside policies and demands as they arose. For example, the FTF reform was able to serve as the district's court ordered desegregation plan; the building reform plans served as the state mandated school improvement plans; and when NCLB was enacted, the district was already in the process of making the improvements needed to meet the mandated annual yearly progress requirements.

Clarity

This element was perhaps the most difficult to achieve, as some aspects of the initiative were sharpened as the initiative moved forward. There were early problems with clarity in the administrative structure, evidenced, for example, by "crossed" lines of reporting for the school improvement facilitators (central office staff assigned to support buildings in the change process); and by several administrators being assigned similar responsibilities, e.g., for professional development. The implementation process itself might be described as achieving ever-improving clarity about the district's purposes and structures for the reform.

Recent studies have made observations about factors effective in creating the clarity needed for meaningful change at the building level. In line with these studies (e.g., Cawelti & Protheroe, 2001), several lessons emerged here:

- Staff need to establish high, clear and fair academic and behavioral standards. Understanding what is expected in terms of the work and the way in which staff relate to other adults and students in the school is clearly linked to good relationships among staff and students. In turn, strong relationships are linked to positive long-term outcomes.
- Not only are high standards a necessary foundational component, but there is a need for a systemic instructional approach. Initially, the district allowed schools to identify their own instructional strategies for improving student performance. Over time, there was a realization that a more coherent, systemic curriculum and instructional approach was needed to provide the supports necessary to create meaningful differences in student outcomes. Increased use of small group formats was implemented and the central office provided instructional coaches in every building to help integrate more effective instructional strategies into every classroom.
- Professional development is a necessary condition for successfully reforming instruction. Early on, the initiative partners all supported the need for system-wide professional development by involving all building staff in Roundtables to learn about the components of FTF. This was followed by the ongoing allocation of resources to professional development – most notably in the weekly early release time that enabled building staff to work

together on self-assessment and strategizing; and through the incorporation of trainings for all teachers on instructional strategies like cooperative learning.

THE RELATIVE IMPORTANCE OF THE REFORM COMPONENTS

There is evidence here that being in the same SLC for two years or more is important for creating stronger relationships between students and teachers, which, in turn, leads to better long-term outcomes. But additional research is needed to identify the optimal period for students to stay within a community, and to further test the critical components of a SLC (e.g., what size is optimal, should they be thematic at the high school level, etc.). What is clear, however, is the central importance of student-teacher relationships to the improvements observed in KCKPS.

There is a general perception in the field that student engagement is the best predictor of long-term outcomes such as attendance and test scores. But results from this study suggest that teacher support is much more sensitive to changes in learning contexts and, in turn, better predicts long-term outcomes than does engagement. This finding demonstrates in a systematic way what is clear in most anecdotal accounts of the vital experience to which most individuals credit their later success – a teacher, whom they can name, who made a critical difference in their lives by the caring and concern they showed.

Interestingly, the best predictor of how well-supported students feel by their teachers is the degree to which they perceive that their teachers have high academic and behavioral standards for them. This suggests that the combination of "challenge" (high standards) and "support" (relationships) are necessary and sufficient conditions for better performance in school. This aligns well with what is known about the general process of human development. Children exhibit the most growth when they are in contexts that provide the necessary combination of pressure – or challenge – and support that enables progress in achieving developmental milestones.

THE ONGOING ROLE OF CONTINUOUS MONITORING IN REFORM

The experience of reform in Kansas City illustrates the importance of having access to accurate information with which to monitor progress. This type of information can be critical to the sustainability of these efforts. The feedback this information provides can both shape adjustments to improvement strategies at both the district and building levels, and fuel staff commitment to reform activities.

A challenge that was highlighted in the Kansas City reform is one that is being, or will be, faced by many districts. In order for districts to guide their efforts with information on students' progress, and to assess building or district improvements over time, data systems are needed that – first and foremost – accurately track critical information for individual students. Many districts collect data at the building level in order to meet reporting requirements, e.g., average daily attendance at a building. But they lack the capacity to store and analyze data in a format that informs them of how many students are actually meeting important benchmarks, e.g., the proportion of students in a building missing no more than one day per month of school. Monitoring the progress of students, rather than managing buildings, is a necessary step in "leaving no child behind." The ability to link and track different types of student information over time (e.g., demographic characteristics, attendance, classroom assignments, test scores, promotion and graduation) will be critical to assessing progress and making the adjustments necessary for continuous improvement.

Further, the ability to demonstrate progress with relevant data can help sustain the ongoing commitment necessary to make long-term reform a reality. Toward the end of the evaluation period in Kansas City, staff interviews revealed a "snowball" effect from feedback in sustaining the desired changes. That is, as staff began to perceive success in longitudinal data, they redoubled their efforts and raised their expectations. This observation is consistent with the quantitative findings that suggest the pace of improvement in outcomes is accelerating. If staff do indeed redouble their efforts when they perceive those efforts to be successful, ongoing data collection is needed to inform those perceptions.

For these reasons, an essential component for improving the effectiveness of schools is the incorporation of systems capable of capturing longitudinal data that monitor and demonstrate the progress being made in both the short and long term.

SUMMARY

Turning the tide in an urban district serving primarily economically disadvantaged youth from a system with disaffected students who do not come to school and who perform poorly to one with students who do attend school regularly, are engaged in their work and develop the necessary skills for a successful transition to adulthood requires committed, clear and consistent leadership; several years to accomplish reforms; teachers working with the same students for longer periods of time; high, clear and fair academic and behavioral standards; teachers who have good relationships with students; effective instructional strategies (e.g., active instruction techniques); teachers who are supported by district leaders; and data to show staff how students are fairing and when improvements are achieved. None of these findings is surprising or new. But the experience of KCKPS and its partners in implementing the FTF reform has demonstrated that this can, in fact, be accomplished and offers an effective pathway for doing so.

REFERENCES

American Institutes for Research. (1999). *An Educator's Guide to Schoolwide Reform*. Arlington, VA: Educational Research Service.

Anderson, S.E. (2003). "The School District Role in Educational Change: A Review of the Literature." *International Centre for Educational Change Working Paper #2*. Toronto: Ontario Institute for Studies in Education.

Battistich, V., Solomon, D., and Kim, D. (1995). "Schools As Communities, Poverty Levels Of Student Populations, And Students' Attitudes, Motives, And Performance." *American Educational Journal*, 32, 627-58.

Berends, M., Chun, J., Shuyler, G., Stockley, S. & Briggs, R.J. (2002). *Challenges of Conflicting School Reforms: Effects of New American Schools in a High-Poverty District*. Santa Monica, CA: Rand.

Berends, M., Kirby, S.N., Naftel, S. & McKelvey, C. (2001). *Implementation and Performance in New American Schools: Three Years into Scale-up*. Santa Monica, CA: Rand. Go to http://www.rand.org/publications/MR/MR1145.

Borman, G.D., Hewes, G.M., Overman, L.T. & Brown, S. (2003). "Comprehensive School Reform and Achievement: A Meta-Analysis." *Review of Educational Research*, 73(2), p.125-230.

Brown, B.V. & Emig, C. (1999). "Prevalence, Patterns, and Outcomes." In D.J. Besharov (ed.). *America's Disconnected Youth: Toward a Preventive Strategy* (pp. 101-116). Washington, DC: CWLA Press.

Cawelti, G. (2003). *What Do We Know: Benchmarking to Eliminate the Achievement Gap.* A Special Report Issued by the Laboratory for Student Success. Philadelphia: The Mid-Atlantic Regional Educational Laboratory at Temple University.

Cawelti, G. & Protheroe, N. (2001). *High Student Achievement: How Six School Districts Changed into High-Performance Systems.* Arlington, VA: Educational Research Service.

Connell, J.P. & Klem, A.M. (2000). "You Can Get There From Here Using A Theory of Change Approach to Plan Urban Education Reform." *Journal of Educational and Psychological Consulting,* 11(1), 93-120.

Connell, J.P., Halpern-Felsher, B., Clifford, E., Crichlow, W., & Usinger, P. (1995). "Hanging In There: Behavioral, Psychological, and Contextual Factors Affecting Whether African-American Adolescents Stay in School." *Journal of Adolescent Research,* 10(1), 41-63.

Connell, J.P., & Wellborn, J.G. (1991). "Competence, Autonomy, and Relatedness: A Motivational Analysis of Self-System Processes." In M.R. Gunnar & L.A. Sroufe (Eds.), Self Processes in Development: *Minnesota Symposium on Child Psychology* (Vol. 23, pp. 43-77). Chicago: University of Chicago Press.

Cook, T.D., Murphy, R.F. & Hunt, H.D. (2000). "Comer's School Development Program in Chicago: A Theory-based Evaluation." *American Educational Research Journal,* 37(2), 535-97.

Croninger, R.G., & Lee, V.E. (2001). "Social Capital and Dropping Out of High Schools: Benefits to At-Risk Students of Teachers' Support and Guidance." *Teachers College Record,* 103(4), 548-581.

Desimone, L. (2002). "How Can Comprehensive School Reform Models be Successfully Implemented?" *Review of Educational Research,* 72(3), 433-479.

Felner, R.D., Jackson, A.W., Kasak, D., Mulhall, P., Brand, S., & Flowers, N. (1997). "The Impact of School Reform for the Middle Years: Longitudinal Study of a Network Engaged in Turning Points-Based Comprehensive School Transformation." *Phi Delta Kappan,* 78, 528-532, 541-550.

Finn, J.D. (1993). *School Engagement and Students At Risk.* Washington, DC: National Center for Education Statistics.

Finn, J.D. & Rock, D.A. (1997). "Academic Success Among Students At Risk for School Failure." *Journal of Applied Psychology,* 82, 221-234.

Gambone, M.A., Klem, A.M., Moore, W.P. & Summers, J.A. (2002). *First Things First: Creating the Conditions and Capacity for Community-wide Reform in an Urban School District.* Philadelphia: Gambone and Associates.

Goodenow, C. (1993). "Classroom Belonging Among Early Adolescent Students: Relationships to Motivation and Achievement." *Journal of Early Adolescence,* 13(1), 21-43.

Granger, R.C. (1998). "Establishing Causality in Evaluations of Comprehensive Community Initiatives." In K. Fulbright-Anderson, A.C. Kubisch & J.P. Connell (eds.). *New Approaches to Evaluating Community Initiatives. Volume 2. Theory, Measurement, and Analysis.* A Publication of the Roundtable on Comprehensive Community Initiatives for Children and Families. Washington, DC: The Aspen Institute.

Halperin, S. (ed.) (1998). *The Forgotten Half Revisited: American Youth and Young Families,* 1988-2008. Washington, DC: American Youth Policy Forum, Inc.

Hauser, R.M. & Sweeney, M.M. (1997). "Does Poverty in Adolescence Affect the Life Chances of High School Graduates?" In Greg J. Duncan & Jeanne Brooks-Gunn (eds.). *Consequences of Growing Up Poor* (pp. 541-595). New York: Russell-Sage.

Institute for Research and Reform in Education. (2003). *First Things First: A Framework for Successful School-site Reform* (V.4). Kansas City, MO: Ewing Marion Kauffman Foundation.

Institute for Research and Reform in Education. (1998). *Research Assessment Package for Schools (RAPS) Manual.* Philadelphia: Institute for Research and Reform in Education.

Kemple, J. & Snipes, J. (2000). Career Academies: Impacts on Students' *Engagement and Performance in High School.* New York: MDRC.

Kronley, R.A. & Handley, C. (2003). *Reforming Relationships: School Districts, External Organizations, and Systemic Change.* Retrieved December 17, 2003 from Annenberg Institute for School Reform at Brown University website: http://www.schoolcommunities.org.

Lee, V.E., & Smith, J.B. (1993). "Effects of School Restructuring on the Achievement and Engagement of Middle-Grade Students." *Sociology of Education,* 66, 164-187.

Lee, V. E. & Smith, J. (1994). "Effects of Restructured Teacher Worklife on Gains in Achievement and Engagement for Early Secondary School Students." Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Lee, V.E., & Smith, J.B. (1995). "Effects of High School Restructuring on Gains in Achievement and Engagement for Early Secondary School Students." *Sociology of Education,* 68, 241-270.

Lee, V.E., & Smith, J.B. (1996). "Collective Responsibility for Learning and its Effects on Gains in Achievement for Early Secondary School Students." *American Journal of Education,* 104(2), 103-147.

Lee, V.E., Smith, J.B., & Croninger, R.G. (1995). "Another Look at High School Restructuring: More Evidence that it Improves Student Achievement, and More Insight into Why." *Issues in Restructuring Schools,* 9, 1-10.

Little, J.W. (1982). "Norms Of Collegiality And Experimentation: Workplace Conditions of School Success." *American Educational Research Journal,* 19, 325-340.

Little, J.W. & McLaughlin, M.W. (Eds.) (1993). *Teachers' Work: Individuals, Colleagues, and Contexts.* New York: Teachers College Press.

Louis, K.S. (1995). "Teacher Engagement and Real Reform in Urban Schools." In B. Williams (Ed.), *Closing the Achievement Gap: A Vision to Guide Change in Beliefs and Practice* (pp. 81-102). Oak Brook, IL: Research for Better Schools and North Central Regional Educational Laboratory.

Louis, K.S. (1992). "Restructuring and the Problem of Teachers' Work." In A. Lieberman (Ed.), *The Changing Contexts of Teaching. Ninety-first Yearbook of the National Society for the Study of Education. Part I.* (p. 138-156). Chicago: The University of Chicago Press.

Maggs, J.L., Frome, P.M., Eccles, J.S. & Barber, B.L. (1997). "Psychosocial Resources, Adolescent Risk Behavior and Young Adult Adjustment: Is Risk Taking More Dangerous for Some Than Others?" *Journal of Adolescence,* 20: 103-119.

Marks, H.M. (2000). "Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years." *American Educational Research Journal,* 37(1), 153-184.

Massell, D. & Goertz, M. (2002). "District Strategies for Building Instructional Capacity." In Hightower, A., Knapp, M., Marsh, J. & McLaughlin, M. (eds.). *School Districts and Instructional Renewal.* New York: Teachers College Press.

McLaughlin, M.W. (1993). "What Matters Most in Teachers' Workplace Context." In J.W. Little & M.W. McLaughlin (Eds.), *Teachers' Work: Individuals, Colleagues, and Contexts.* New York: Teachers College Press.

McLaughlin, M. & Talbert, J. (2003). *Reforming Districts: How Districts Support School Reform.* A Research Report. Seattle, WA: Center for the Study of Teaching and Policy at the University of Washington.

National Research Council & the Institute of Medicine. (2004). *Engaging Schools: Fostering High School Students' Motivation to Learn.* Committee on Increasing High School Students' Engagement and Motivation to Learn. Board on Children, Youth, and Families, Division of Behavioral and Social Science and Education. Washington, DC: The National Academies Press.

Northwest Regional Educational Laboratory and National Clearinghouse for Comprehensive School Reform. (2004). *The Catalog of School Reform Models.* Portland, OR: Northwest Regional Educational Laboratory and National Clearinghouse for Comprehensive School Reform.

Olson, L. (2000). "Annenberg Task Force Will Take a New Look at Districts." *Education Week.* Retrieved December 17, 2003 from http://www.edweek.org/ew/ewstory.cfm?slug=25district.h19&keywords=Annenberg%20Institute.

Padilla, Y.C. (1997). "Determinants of Hispanic Poverty in the Course of the Transition to Adulthood." *Hispanic Journal of Behavioral Sciences,* 19(4):416-432.

Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research* (3rd ed.). Orlando, FL: Harcourt Brace.

Plank S.B. & Jordan, W.J. (1997). *Reducing Talent Loss: The Impact of Information, Guidance and Actions on Postsecondary Enrollment* (Report No. 9). Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed At Risk.

Porter, A.C. (1994). "National Standards and School Improvement in the 1990s: Issues and Promise." *American Journal of Education,* 102, 421-449.

Roderick, M. & Engle, M. (2001). "The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing." *Educational Evaluation and Policy Analysis,* 23(3), 197-227.

Rosenholtz, S.J. (1989). Teachers' Workplace: *The Social Organization of Schools.* New York: Longman.

Ryan, A.M. & Patrick, H. (2001). "The Classroom Social Environment and Changes in Adolescents' Motivation and Engagement During Middle School." *American Educational Research Journal,* 38(2), 437-460.

Shouse, R.C. (1996). "Academic Press and Sense of Community: Conflict, Congruence, and Implications for Student Achievement." *Social Psychology of Education,* 1(1) , 47-68.

Skinner, E.A., & Belmont, M.J. (1993). "Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement Across the School Year." *Journal of Educational Psychology,* 85(4), 571-581.

Skrla, L., Scheurich, J. & Johnson, J., Jr. (2000). *Equity-driven Achievement-focused School Districts: A Report of Systemic School Success in Four Texas School Districts Serving Diverse Student Populations.* Austin: The Charles A. Dana Center, The University of Texas at Austin. Retrieved on December 17, 2003 from http://www.utdanacenter.org/research/reports/equitydistricts.pdf.

Slavin, R. (2003). "Converging Reforms." *Education Week.* Retrieved on March 24, 2004 from http://www.edweek.org.

Slavin, R. & Madden, N. (2004). *Scaling Up Success For All:* The First 16 Years. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Snipes, J. Doolittle, F. & Herlihy, C. (2002). *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement.* New York: MDRC for the Council of the Great City Schools.

Solomon, D., Battistich, V., Watson, M., Schaps, E. & Lewis, C. (2000). "A Six-District Study of Educational Change: Direct and Mediated Effects of the Child Development Project." *Social Psychology of Education,* 4, 3-51.

Stallings, J.A. (1977). *Learning to Look: A Handbook on Classroom Observation and Teaching Model.* Belmont, CA: Wadsworth.

Steinberg, L.D., Brown, B.B. & Dornbusch, S.M. (1996). *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do.* New York: Simon and Schuster.

Togneri, W. & Anderson, S.E. (2003). *Beyond Islands of Excellence: What Districts Can Do to Improve Instruction and Achievement in All Schools.* Washington, D.C.: The Learning First Alliance and the Association for Supervision and Curriculum Development.

Viadero, D. (2004). "Call for 'Scientifically Based' Programs Debated." *Education Week.* Retrieved on March 24, 2004 from http://www.edweek.org.

Voelkl, K.E. (1995). “School Warmth, Student Participation, and Achievement.” *Journal of Experimental Education,* 63, 127-138.

Weiss, Carol H. (1995). “Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families.” In J.P. Connell, A.C. Kubisch, L B. Schorr & C.H. Weiss (eds.). *New Approaches To Evaluating Community Initiatives. Volume 1. Concepts, Methods, and Contexts.* A publication of the Roundtable on Comprehensive Community Initiatives for Children and Families. Washington, DC: The Aspen Institute.

Willingham, W.W., Pollack, J.M., & Lewis, C. (2002). “Grades and Test Scores: Accounting for Observed Differences.” *Journal of Educational Measurement,* 39(1), 1-37.

Wohlstetter, P., Van Kirk, A.N., Robertson, P.J., & Mohrman, S.A. (1997). *Organizing for Successful School-Based Management.* Alexandria, VA: Association for Supervision and Curriculum Development.

APPENDIX A: Technical Appendix

This appendix presents information about:

- measures used in the study;
- models used to estimate impact and relationships; and
- statistics used to evaluate the relationships between the variables in the FTF framework.

First, we present a summary of the data collection process, including how the intermediate outcome and implementation variables were created, and the psychometric properties of each scale. Second, the appendix provides a discussion of how we set high-risk and optimal thresholds on the implementation measures, as well as the intermediate and long-term outcomes. Third, this appendix describes the administration of the survey, observation and qualitative measures. Finally, the appendix provides an overview of our data analysis strategy for estimating the relationships between First Things First implementation and student and staff outcomes, including a description of the models and the particular statistics used to make inferences.

DATA COLLECTION

The data for the study were collected over six consecutive school years, beginning in the 1997-1998 school year (planning and pre-implementation for the first cluster) through the 2002-2003 school year. Data were collected using a combination of methods, including student and staff surveys, classroom observations, student records and assessment data, and qualitative ratings. Table A-1 shows the data collection schedule for each method across the study years.

Table A-1: Data Collection Schedule

Clusters	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03
Student Level Measures						
Student survey Intermediate Outcomes	✔	✔	✔	✔	✔	✔
Student Survey Implementation Measures	✔	✔	✔	✔	✔	
Attendance	✔	✔	✔	✔	✔	✔
State Math Test Scores			Pilot Year	✔	✔	✔
State Reading Test Scores			Pilot Year	✔	✔	✔
Cohort Graduation Data				✔	✔	✔
High School Dropout				✔	✔	✔
Building Level Measures						
Staff Survey Intermediate Outcomes of Engagement and Support Measures	✔	✔	✔	✔	✔	
Staff Survey Perceptions of Implementation Measures	✔	✔	✔	✔	✔	
Classroom Observations		✔ WY and WA only*	✔ WY and WA only	✔	✔	
Small Learning Community Membership (Student records)			✔ WY only	✔ WY and WA only	✔	✔

*WY=Wyandotte Cluster WA=Washington Cluster

SURVEY ADMINISTRATION

STAFF SURVEY

The Staff Survey was designed by the Research Management Team[29] and includes items that tap into three constructs discussed in this report: perception of critical features, staff engagement, and staff's experience of support from colleagues and system leaders. The survey also includes basic demographic questions, including current position, duration of current position and other work history items, gender, and ethnicity.

The Staff Survey was administered in all clusters each spring from 1998 to 2002. All educational staff members at each school were asked to complete this survey during staff meetings. The response rates across years and clusters to the staff survey ranged from 80 to 100 percent. Tables showing the sample sizes and response rates for the staff survey by cluster and year can be found in the Technical Report, available at www.ydsi.org.

In 2000, several items on the Staff Survey were revised or deleted prior to the 2001 administration. The early outcome items were removed from the survey, and items asking about staff perceptions of individual critical features were either deleted or revised to ask about experiences within their SLC rather than in their school.[30] These variations in wording led to several of these items being dropped from the analyses. Items that remained consistent were retained in the analyses. The Staff Survey was discontinued by the district after the 2001-2002 school year.

STUDENT SURVEY

The Student Survey consisted of a revised version of the Research Assessment Package for Schools Student Version (RAPS-S), developed by the Institute for Research and Reform in Education (IRRE) and assesses student engagement, teacher support and students' perceptions of the four student critical features. There are two versions of the survey: one for elementary students and another for secondary students (middle and high school). The elementary survey was administered to students in grades 3 to 5; the secondary survey to grades 6 to 12.

The Student Survey was administered each spring from 1998 to 2002. A shortened version of the survey was administered, beginning in Spring 2003, and included only engagement and teacher support questions. The survey was administered in a group format to students in classrooms during regular class periods with an adult reading each question aloud in order to minimize literacy differences. Spanish and other non-English language versions of the survey were administered to non-English speaking students by their ESL teachers. Surveys were usually administered to special education students separately in order to provide them with extra time as needed. The response rates across years ranged from 85 to 99 percent for elementary, 80 to 100 percent for middle and 54 to 96 percent for high school students. Tables showing the sample sizes and response rates by cluster and year can be found in the Technical Report, available at www.ydsi.org.

CLASSROOM OBSERVATIONS

The observation system developed for the evaluation of First Things First (FTF) was designed to provide information about the implementation progress of the critical features at the classroom level. In particular, the protocol captured the extent of implementation of critical feature #4 (Provide enriched and diverse opportunities to learn, perform and be recognized). In addition, information was collected within the observation system about two other critical features:

- Set high, clear and fair academic and behavioral standards, and
- Lower student-adult ratios

Classroom observations focused on documenting the occurrence and frequency of use of specific instructional activities (e.g., lecture, cooperative learning, silent reading), grouping structures (e.g., whole class, small group, individuals), prevalence of specific teacher behaviors (e.g., leading instruction, observing, waiting) and interactions with students (e.g., teacher directs question to individual student, teacher responds to individual student's question). The observation protocol also captured information about the physical environment of the classroom (e.g., desks in circles, performance standards posted in room, computer in room). The observation system

[29] This survey was adapted from measures originally used by IRRE.

[30] These changes were made to accommodate the initiative leaderships' desire to have data that more accurately reflected the direction of implementation, so the information could inform their ongoing decisions.

was adapted with permission from the developer of the approach (Stallings, 1977).

Trained and experienced educators (or professionals with years of classroom observation experience) were hired specifically to conduct classroom observations for this study. Observers were not employees of the district nor were they employed by the district to conduct the observations. All observers participated in training experiences using written and videotaped classroom vignettes of actual classroom activity and group discussions of coding choices to align coding decisions across observers with known classroom practices and activities. Observers who did not reach a pre-established criterion for intra-rater and inter-rater reliability were not allowed to conduct observations.

After one year of protocol development and pilot-testing (1997-1998), in which 109 classrooms were used to inform revisions to the procedures and protocol, a baseline was established in 1998. For the second cluster of schools to enter FTF implementation, a revision of the protocol was again made to address the long-term usability and sustainability of the protocol for internal monitoring purposes. This second version was pilot-tested during Fall 2000.

Observations used in the research reported in this document were conducted in 1,058 elementary, middle and high school classrooms during the spring terms of the 1998-1999 to 2001-2002 school years for the first two clusters of schools to implement FTF. Later, observations were conducted in the remaining two clusters of schools (460 classrooms) using a modified version of the protocol in Spring 2001 and 2002. Observations were conducted in kindergarten through twelfth grade classrooms during reading/language arts, mathematics, science and social studies instruction.

STUDENT RECORDS AND ASSESSMENT DATA

Student records data were obtained from the district research office annually, and included enrollment and attendance data, dropout and graduation codes, student demographic information, and building and SLC membership. All data were cleaned within and across years to the extent possible. The research team was unable to obtain accurate student level data prior to the 1997-1998 school year; all data prior to that year was incomplete or contained many inaccuracies that could not be corrected. Student test scores were obtained from district data sets and from the Kansas State Department of Education.

QUALITATIVE DATA COLLECTION

A full description of the methods used in the qualitative data collection can be found in the Technical Report.

Creating Implementation and Intermediate Outcome Measures

Scales were created from student and staff surveys, observations, school data, and qualitative data to measure the intermediate outcomes and the implementation measures of FTF. Several steps were taken to develop the scales (see the Technical Report at www.ydsi.org for complete analytic description and results).

1. Existing scales were evaluated to determine their degree of statistical fit for one factor using confirmatory factor analysis techniques. In the confirmatory factor analysis, all items were fixed to load on a single factor and the degree of fit and factor loadings were estimated.

2. For scales that did not demonstrate adequate fit using a one-factor model, exploratory factor analyses were conducted to determine the appropriate number of factors for the scale. The resulting factor structures were analyzed using confirmatory factor analysis to assess degree of fit.

3. Reliability estimates were calculated for all scales using a method of structural equation modeling to estimate the true and error variance.

4. Correlations were calculated between all items and their own scale (convergent correlations) and with other scales (divergent correlations). To the degree that the convergent correlations were higher than the divergent correlations, the s cale scores were evaluated to have adequate construct validity.

Survey Measures

Student surveys were used to construct measures of student intermediate outcomes; teacher support and engagement; and student experiences of active and

connected learning, academic and behavioral standards, and continuity of care. Staff surveys were used to assess staff intermediate outcomes of support and engagement, and perceptions of critical features implementation. Table A-2 shows the scales associated with each of these measures.

OBSERVATIONAL MEASURES OF INSTRUCTION AND STRUCTURE

Classroom observations were used to create the measures for small grouping, active learning and classroom ratios. Small grouping and active learning were computed by calculating the proportion of observation cycles in the classroom where small grouping structures and active learning strategies were in use. Threshold levels were set on these average proportions. Classroom ratios were calculated by dividing the number of students enrolled in the class by the number of adults present.

SLC stability was created from student records data. A student was coded as being experiencing low SLC stability if s/he was assigned to an SLC for less than two years. Conversely, a student was coded as experiencing high SLC stability if s/he attended SLCs for two or more years.

Table A-2: Implementation of Intermediate Outcome Measures and Scale Reliabilities

Measure	Sample Items	Scale Reliability[a]	
		E	S
	Intermediate Outcomes		
Teacher Support (student-7 items)	• My teachers care about how I do in school. • My teachers don't seem to have enough time for me.	.51	.64
Student Engagement (student-8 items)	• I work very hard on my schoolwork • I often come to class unprepared	.68	.69
Staff Individual Engagement (staff-12 items)	• My job has become just a matter of putting in time • I look forward to going to work in the morning.	.80	
Collective Engagement and Colleague Support (staff-6 items)	• Staff at this school do what is necessary to get the job done right. • Staff in this school go out of their way to help each other	.87	
System Support (staff-10 items)	• Staff get the resources they need from the central office to support work they do with students. • School building administrators support staff making their own decisions about their students.	.88	
	Implementation Measures		
Active and Connected Learning (students-12 items)	• Students get to work on projects that they help design • Students talk about connections between our work in school and what is going on in our lives outside of school	.74	.81
Academic and Behavioral Standards (students-10 items)	• Teachers show us examples of the kinds of work that can earn us good grades • All adults in this school treat all students the same when it comes to following the rules	.70	.75
Critical Features Implementation (staff-12 items)	• To what extent is this critical feature [lower student-adult ratios] being implemented now in your SLC? • To what extent is this critical feature [give school staff more instructional autonomy] being implemented now in your SLC	.84	

[a] Scale reliabilities were computed for each year. Table A-2 reports the average reliability across all years. In all cases, the scale reliabilities varied no more than .04 across years.

Thresholds for Measures

All analyses assessing the relationships among the variables in the FTF framework were conducted on categorical variables representing risk, indeterminate (or moderate) and optimal levels on each variable. To create these levels (or thresholds), cutpoints were set on all measures. For measures that had been previously used in other work (e.g., Bridges, 2001), existing cutpoints were used to set thresholds.

The end result of these analyses was a set of categorical measures that defined groups of individuals as either optimal (in good shape with respect to this element or sub-element of the framework) or at risk (having difficulty in this area).

Staff and Student Survey Cutpoints

The following staff and student survey constructs were answered using a 4-point Likert Scale ranging from 1 (Not At All True) to 4 (Very True):

- Students: Engagement, Teacher Support, Active and Connected Learning, and Academic and Behavioral Standards.
- Staff: Engagement, Support from Colleagues, and System Support.

The threshold cutpoints for these outcomes are included in the body of the report.

Implementation Cutpoints

Cutpoints for the staff survey are based on a 5-point scale from 1 (Implemented for No Students) to 5 (Implemented for All Students). Cutpoints for the student survey are based on a 4-point Likert scale from 1 (Not At All True) to 4 (Very True) and are also included in the body of the report.

Analytic Strategy

The analytic strategy for assessing changes associated with FTF implementation and subsequent intermediate and long-term outcomes involved several steps. Typically, when both building and student/staff data were available, the most efficient and comprehensive approach to assessing the effects of building characteristics on changes in students and staff was to utilize a three-level hierarchical model, where changes in students and staff were modeled over time as a function of both individual and building-level predictors. However, the data structure available precluded answering all of the evaluation questions in a single set of analyses.[31]

The analytic strategy, therefore, relied on constructing a logical argument for the relationship between FTF implementation and student and staff outcomes. The steps in building the evidence were:

- First, we conducted analyses to determine whether measures of implementation showed improvement associated with each additional year of implementation. This addressed the question of whether more implementation of FTF was associated with greater degrees of implementation of the critical features.
- Second, we conducted analyses to examine changes in student and staff outcomes over the course of implementation. This addressed the question of whether more implementation was associated with better outcomes.
- Third, we examined how specific building contextual factors (e.g., leadership, professional development, decision-making) were related to levels of intermediate outcomes and implementation of critical features. In addition, we examined how levels of critical feature implementation were related to levels of intermediate and long-term outcomes for students. This addressed the question of whether the links hypothesized in the reform model between what happened in buildings and the experiences of staff and students were related to improved outcomes.
- Finally, we conducted analyses of the state test score data to compare progress for Kansas City with the state in reading and math.[32]

To the degree that each link in the argument provides positive evidence, the whole allows us to begin to make attributions about FTF and its contribution to improving outcomes.

[31] Because implementation was phased in over three years, the number of years each cluster is implementing FTF varies. Also, survey and observation data collection were stopped one year prior to the end of the study. Finally, limited cohorts of graduation, dropout data prior to 1997-1998 were not available due to inaccurate records; and state test scores were not available prior to 2000-2001 when the state test was substantially revised.

[32] Because there is no baseline data available to do comparisons of FTF implementation versus no FTF implementation (the new form of the state test was started in the 2000-2001 school year), we compared the district results to the state to rule out history or state-wide policy effects on changes in reading and math scores.

All analyses, except where otherwise noted, were conducted using data from all four clusters and their comprehensive high schools (we excluded data from the alternative school program and magnet high school). In addition, all analyses utilized a logistic modeling framework to estimate the effects of categorical predictors on dichotomous outcome variables, and to examine the changes in proportions of student, staff and implementation measures in optimal and risk categories on the dependent variables.

DEPENDENT MEASURES

For all research questions, we examined both the changes in the optimal category and in the risk category of particular variables. The dependent variables, therefore, were dichotomous variables, where a 1 represented either optimal or risk status (depending on the analysis) and a 0 represented being not optimal or not risk. In the case of graduation, promotion, retention and dropout, single analyses were conducted. For these dichotomous dependent variables, a 1 represented the occurrence of the event (e.g., the student graduated or dropped out of school) and a 0 represented non-occurrence of the event.

EXPLORATORY ANALYSES

Exploratory, unconditional, two-level logistic models were estimated to determine if the implementation and intermediate and long-term outcomes had sufficient variability across buildings and years to use a hierarchical modeling structure to analyze the data. The Level-1 and Level-2 models for these analyses were:

$\eta_{ij} = \beta_{0j} + \varepsilon_{ij}$ (Level-1) where η_{ij} is the estimated probability for a given student in a given building and year, β_{0j} is the average proportion across all students for a given building and year (group mean), and ε_{ij} is the random error associated with each student.

$\beta_{0j} = \gamma_{00} + u_{0j}$ (Level-2) where β_{0j} is the average proportion across all students for a given building and year, g00 is the average proportion across all buildings and years (grand mean), and u_{0j} is the unique variance associated with each building and year.

A detailed description of these exploratory analyses can be found in the Technical Report at www.ydsi.org.

ASSESSING STRENGTH OF IMPACT EFFECTS

The typical measures of effect size in logistic models include odds-ratios and relative risk values. The odds-ratio is the exponentiated coefficient from the logistic regression analysis. The relative risk is the ratio of frequencies for an outcome under two different conditions for example, under the baseline condition of no FTF implementation, and under the exposure condition of multiple years of FTF implementation.

To assess the strength of individual statistical effects of implementation, we converted the odds-ratios estimated in the models to relative risks. Relative risks were chosen to represent the magnitude of effect rather than odds-ratios because they are relatively straightforward to interpret. In addition, odds-ratios tend to overestimate the effect of the intervention, particularly when the initial risk is quite high and the effect is large. The formula for converting odds-ratios into relative risk is the following:

$RR = OR/((1 - p_0) + (p_0 * OR))$ where

OR = estimated odds ratio and
p_0 = the proportion of the outcome for the comparison group.

For the effects of implementation year, the comparison group is the baseline year (or Year 1 for non-baseline analyses). For the effects of intermediate outcomes, the comparison group is the indeterminate group (or the average student); and for the effects of implementation, the comparison group is the low implementation group.

Relative risks are interpreted as the change in likelihood of a given outcome, given two different conditions (e.g., the change in likelihood that students will score proficient or above on the state assessment when in high levels of FTF implementation instead of low levels of FTF implementation). Positive relative risk values range from 1.00 to positive infinity and negative relative risk values range from .999 to .000. A relative risk value of 1.5 can be interpreted as a 50

percent increase in the likelihood of an event occurring or, alternatively, as a student being 1.5 times more likely to experience the outcome. Negative risks can be interpreted as the decrease in likelihood of an outcome occurring. For example, a relative risk value of .80 represents a 20 percent decrease in the likelihood of an outcome occurring (1 - .80).

ANALYSES TO ASSESS CHANGE IN IMPLEMENTATION AND OUTCOMES OVER TIME

The analyses to assess changes in implementation and outcomes over time followed the same general strategy. Analyses were conducted for each student and staff outcome, as well as for each implementation measure described above, both for optimal levels and high-risk levels. Two-level logistic models were estimated for each dichotomous outcome variable, with the level of predictors in the model dependent upon the question to be answered. All predictors were transformed into categorical thresholds representing high, moderate and low levels of each predictor. The covariates were coded as 1 or 0.[33] At the student level of the models, we controlled for population differences by co-varying out the effects of student demographic characteristics of ethnic status, mobility, SES and gender. All student-level covariates were centered around their grand mean, as typically done in analysis of covariance (Pedhazer, 1997). At the building level of the model, we controlled for the differences due to cluster membership, since the clusters implemented FTF across the span of three years. Cluster membership was coded as a set of dummy vectors, with Wyandotte as the comparison cluster.

Student-level outcomes and implementation measures, with a random building/year intercept

In these models, student-level control variables and a building-level predictor of implementation year was used to estimate the change in student perceptions of teacher support and engagement; perceptions of active and connected learning, and academic and behavioral standards; student membership in SLCs; and student long-term outcomes (e.g., test scores and attendance). The general Level-1 model for these analyses is shown below:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{RACE}) + \beta_{2j}(\text{SES}) + \beta_{3j}(\text{GENDER}) + \beta_{4j}(\text{MOBILITY}) + \varepsilon_{ij}$$

where η_{ij} is the estimated probability for a given student in a given building and year, β_{0j} is the average proportion across all students for a given building and year (group mean), β_{1j} to β_{4j} are the average effects of the covariates on the student outcome across all students for a given building and year, and ε_{ij} is the random error associated with each student. RACE, SES, GENDER and MOBILITY are dummy-coded variables with 1 representing minority status, free and reduced-price lunch status, males, and attendance at more than one school, respectively.

In these models, only the intercept (average outcome across buildings and years) was allowed to vary. All coefficients were fixed to be the same value across buildings (for more detailed analysis of logic for fixed coefficients, see the Technical Report at www.ydsi.org). Subsequently, the general Level-2 model for these analyses is shown below:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{IMPYEAR}) + \gamma_{02}(\text{CLUSTER}) + u_{0j}$$

where IMPYEAR is equal to the number of years of implementation of FTF (0 = baseline to 4 = four years of implementation) for a building within a given school year,

CLUSTER is the effect of cluster membership of the building (with Wyandotte cluster coded as the comparison group),

γ_{00} is the mean of the dependent variable across buildings and years,

γ_{01} represents the effect of IMPYEAR

γ_{02} represents the effect of CLUSTER.

u_{0j} represents the degree of variability in the mean of the dependent variable across buildings and years.

[33] Gender was coded 1 for males, SES was coded 1 for free and reduced-price lunch, mobility was coded as 1 for attended more than one school, and ethnicity was coded 1 for minority status. In analyses examining the interaction effects of ethnicity or SES, ethnicity was coded 0 for white students, 1 for African-American students, and 2 for Hispanic students.

A second set of analyses were conducted to determine if there were any interactions between the length of time in implementation and minority or SES status in predicting the student outcomes. These cross-level models had the same Level-1 model, but allowed the effects of ethnicity or SES to vary randomly across buildings and years. The resulting Level-2 models for these analyses were the following:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{IMPYEAR}) + \gamma_{02}(\text{CLUSTER}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{IMPYEAR}) + u_{1j} \quad \text{(RACE) OR}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{IMPYEAR}) + u_{2j} \quad \text{(SES)}$$

The effects of IMPYEAR on b1j or b2j represent the differential effect of RACE or SES, respectively, on the student outcomes across implementation years. To follow-up significant RACE or SES interactions, separate two-level logistic analyses were conducted for each subgroup, using the strategy described above (omitting the covariate SES or RACE, depending on the subgroups to be analyzed).

CHANGES IN GRADUATION AND DROPOUT RATE

The analyses for graduation and dropout were conducted using different analytical models than the other outcomes. Because the data for graduation and dropout rates were collected for three longitudinal cohorts of students (the graduating classes of 2000-2001, 2001-2002 and 2002-2003), we conducted Cox regression proportional hazards survival analyses to determine if the dropout and graduation rates were changing across the three cohorts. Survival analysis allows us to analyze event data by censoring cases that do not exhibit the particular event (e.g., dropping out or graduating). It also allows us to examine the changes in the dropout and graduation rates longitudinally.

The survival analyses included the student covariates described above in the previous section, as well as a time variable (total days enrolled in high school that has a maximum of 718 days – the maximum time for graduation or dropout to occur over a four-year high school career) and a cohort effect (representing the change from year to year). The analyses were conducted separately for each of the four high schools because of the confounding of cohorts and implementation years. The general survival model is the following (single-level):

$$H(t) = H_0(t) \times \exp(b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k)$$

where $X_1 \ldots X_k$ are a collection of predictor variables and $H_0(t)$ is the baseline hazard at time t, representing the hazard for a person with the value 0 for all the predictor variables. This model can also be expressed as a hazard ratio, or the ratio of baseline hazard divided by the end of time hazard. By dividing both sides of the above equation by $H_0(t)$ and taking logarithms, we obtain:

$$\ln\left[\frac{H(t)}{H_0(t)}\right] = b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

We call $H(t) / H_0(t)$ the hazard ratio. The coefficients $b_i \ldots b_k$ are the effects of the covariates and predictors estimated by Cox regression, and can be interpreted in a similar manner to that of multiple logistic regression. For dropout, to the degree that the hazard functions for later cohorts are less than the baseline hazard function, there is a decrease in the dropout rate. Conversely, for graduation, to the degree that the hazard functions are greater than the baseline hazard rate, there is an increase in the graduation rate.

Changes in staff and non-student implementation measures

To estimate changes in staff outcomes (staff engagement, system support, and colleague support and engagement) and perceptions of implementation, and to assess changes in the classroom-level observation measures (active learning, small grouping and class ratios), a similar set of models to the changes in student outcomes was used. However, for these analyses, no staff or classroom characteristics were used as Level-1 covariates. The general Level-1 model for these analyses is the same as the model described in the following:

$$\eta_{ij} = \beta_{0j} + \varepsilon_{ij}$$

where η_{ij} is the estimated probability for a given student in a given building and year,

β_{0j} is the average proportion across all students for a given building and year (group mean), and

ε_{ij} is the random error associated with each student

The Level-2 model was identical to the Level-2 model for changes in student outcomes:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{IMPYEAR}) + \gamma_{02}(\text{CLUSTER}) + u_{0j}$$

ESTIMATING STATISTICAL EFFECTS OF INTERMEDIATE OUTCOMES AND BUILDING IMPLEMENTATION MEASURES ON STUDENT INTERMEDIATE AND LONG-TERM OUTCOMES

When testing the relationship between variables in the FTF framework, two types of two-level logistic models were used.

Individual student variables relationships with student outcomes

First, individual student-level logistic models were estimated that examine the relationships between the student intermediate outcomes and students' implementation experiences. These student-level models also took into account the student demographic characteristics of gender, race, mobility and SES. These models looked at whether students scoring higher on certain variables (e.g., higher standards) also scored higher on other variables (e.g., more teacher support).

Table A-8 shows the Level-1 models for each of the student-level relationship models.

Building-level implementation relationships with student outcomes

Second, to examine the effects of building-level implementation measures (e.g. classroom observation measures or staff survey measures) on student long-term and intermediate outcomes, different two-level logistic models were estimated. These two-level logistic models enabled us to examine the statistical effect of building-level measures such as the student-teacher ratios observed during core instruction on students' experience of support from their teachers, while taking into account such student

Table A-8: Level-1 Models for Implementation and Intermediate Outcome Statistical Effects on Teacher Support, Engagement, Attendance, and Reading and Math Achievement

Outcome	Level-1 Model
Teacher Support	$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{RACE}) + \beta_{2j}(\text{SES}) + \beta_{3j}(\text{GENDER}) + \beta_{4j}(\text{MOBILITY}) + \beta_{5jj}(\text{ACTIVE}) + \beta_{5j}(\text{STANDARDS}) + \varepsilon_{ij}$
Engagement	$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{RACE}) + \beta_{2j}(\text{SES}) + \beta_{3j}(\text{GENDER}) + \beta_{4j}(\text{MOBILITY}) + \beta_{5j}(\text{TCHSUP}) + \beta_{6j}(\text{PARSUP}) + \beta_{7j}(\text{ACTIVE}) + \beta_{8j}(\text{STANDARDS}) + \varepsilon_{ij}$
Attendance and State Reading and Math Achievement	$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{RACE}) + \beta_{2j}(\text{SES}) + \beta_{3j}(\text{GENDER}) + \beta_{4j}(\text{MOBILITY}) + \beta_{5j}(\text{TCHSUP}) + \beta_{6j}(\text{PARSUP}) + \beta_{7j}(\text{ENGAGEMENT}) + \varepsilon_{ij}$

The Level-2 model was $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{CLUSTER}) + u_{0j}$

Table A-9: Level-2 Models for Implementation and Intermediate Outcome Statistical Effects on Teacher Support, Engagement, Attendance, and Reading and Math Achievement

Set of Building Implementation Measures	Level-2 Model
Staff Survey Measures	$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{CLUSTER}) + \gamma_{02}(\text{STAFFENG}) + \gamma_{03}(\text{SYSSUP}) + \gamma_{04}(\text{COLLSUP}) + \gamma_{05}(\text{CFIMP}) + u_{0j}$
Active Learning and Small Grouping[a]	$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{CLUSTER}) + \gamma_{02}(\text{ACTIVELRN}) + \gamma_{03}(\text{SMALLGP}) + u_{0j}$
Class Ratios and SLC Stability	$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{CLUSTER}) + \gamma_{02}(\text{RATIOS}) + \gamma_{03}(\text{SLCSTAB}) + u_{0j}$

[a] Instruction and small grouping building relationship models were conducted on Wyandotte and Washington clusters only, and on grades 3 to 12 of the observation measures (corresponding to the grades to which the surveys were administered).

characteristics as the extent to which individual students experienced active instruction or high standards. These models allowed us to ask whether buildings scoring higher on certain variables (e.g., classes had lower student-adult ratios) also had improved scores on other variables (e.g., students were less disengaged from school).

Three separate building models were estimated, each of which controlled for student-level variables but also contained a unique set of independent variables. The first building-level model contained structure independent variables (observed ratios and SLC stability), the second included instruction-related independent variables (observed active learning and small grouping formats) and the third contained staff intermediate outcomes (colleague support, system support and staff engagement). We were not able to combine the models estimating the effects of implementation year with the actual measures of implementation because of missing data points (see the Technical Report at www.ydsi.org for a description of the missing data).

The Level-1 models for these analyses are the same as those shown in Table A-8. The Level-2 models for these analyses are shown in Table A-9.

ESTIMATING THE STATISTICAL EFFECTS OF BUILDING CHARACTERISTICS ON IMPLEMENTATION AND INTERMEDIATE OUTCOMES

Univariate chi-square analyses were conducted to estimate the relationships between the qualitatively derived ratings of the building factors (e.g., leadership, participative decision-making, professional development and relationships in the building), the measures of implementation, and staff and student intermediate outcomes. All analyses were conducted on Wyandotte and Washington clusters, as those were the two clusters in which the qualitative data were collected.

A set of univariate chi-square analyses was performed for each dependent variable. The level of significance was set at p = .05 for the set of analyses, and the Type I error rate was controlled for across the analyses within a set using the Holm's modified Bonferroni procedure.[34] Relative risk statistics were computed to determine the change in relative risk for optimal outcomes in low and high implementation, and for high-risk outcomes in low and high implementation. Relative risk values were computed by dividing the frequency of respondents in the high implementation category by the frequency of responses in the low implementation category.

COMPARING STATE ASSESSMENT PERFORMANCE DIFFERENCES BETWEEN KCK AND THE STATE

Two-level logistic models were conducted to assess the degree to which KCK made differential progress on the state assessment relative to the rest of the state. Analyses were conducted for both the change in the proportion of students scoring at the proficient or above level, and for the change in the proportion of students at the unsatisfactory level. Separate analyses were conducted for each grade and subject.

The Level-1 predicts score for all students controlling for student demographics:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}(\text{RACE}) + \beta_{2j}(\text{SES}) + \beta_{3j}(\text{GENDER}) + \varepsilon_{ij}$$

The Level-2 model for these analyses included a dummy-coded vector (0 = KCK, and 1 = State) to represent the state and KCK; a dummy-coded vector TIME to represent the three years of testing (with Year 1 as the comparison group) since the beginning of the new test administration; and a set of interaction vectors representing the interaction between TIME and DISTRICT. To the degree that the interaction is significant, KCK and the state are making progress on the state assessment differentially. The general Level-2 model for this analysis is as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{DISTRICT}) + \gamma_{02}(\text{TIME}) + g\gamma_{03}(\text{DISTRICT*TIME}) + u_{0j}$$

[34] The Holms modification adjusts the Type I error rate across the number of comparisons in a set. The comparison p-values are ordered from smallest to largest. For the first comparison, the p-value is adjusted by the total number of comparisons. If the comparison is significant using the adjusted p-value, the p-value for the next comparison is adjusted by the number of remaining comparisons in the set (n-1 comparisons). The comparison procedure continues until a non-significant result is obtained or all comparisons have been tested.

To the degree that the DISTRICT*TIME interaction was significant, follow-up analyses were conducted for KCK and the state separately to estimate the effects of time on their changes in student performance. The Level-1 model remained the same, and the follow-up Level-2 model is:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TIME}) + u_{0j}$$

Multivariate hypothesis tests were conducted to assess specific changes over time for both the state and KCK separately.

A second set of analyses was conducted to determine if there were any interactions between the Level-2 predictors of DISTRICT, TIME and DISTRICT*TIME with minority or SES status in predicting the student test performance. These cross-level models had the same Level-1 model, but allowed the effects of ethnicity or SES to vary randomly across buildings and years. The resulting Level-2 models for these analyses were the following:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{DISTRICT}) + \gamma_{02}(\text{TIME}) + \gamma_{03}(\text{DISTRICT*TIME}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{DISTRICT}) + \gamma_{12}(\text{TIME}) + \gamma_{13}(\text{DISTRICT*TIME}) + u_{1j}$$

OR

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{DISTRICT}) + \gamma_{22}(\text{TIME}) + \gamma_{23}(\text{DISTRICT*TIME}) + u_{2j}$$

The effects of DISTRICT on β_{1j} or β_{2j} represent the average differential effect of race or SES on the student performance between KCK and the state. The effects of TIME on β_{1j} or β_{2j} represent the average differential effect of race or SES on the student performance across years, holding constant the effect of DISTRICT. The effects of DISTRICT*TIME on β_{1j} or β_{2j} represents a three-way interaction assessing whether the change in effects of SES or RACE on student performance across years varies as function of the differential effects of DISTRICT and TIME, or whether the gap between racial groups or economic groups is changing differentially for state and KCK students. To follow-up significant effects of DISTRICT*TIME, separate two-level logistic analyses were conducted for each subgroup, using the strategy described above (omitting the covariate SES or race, depending on the subgroups to be analyzed). Multivariate hypothesis tests were conducted to assess specific changes over time for both the state and KCK separately.

REFERENCES

Bridges, L. (2001). *Measurement Report for First Things First Evaluations.* Technical report prepared for MDRC. New York: MDRC.

Stallings, J.A. (1977). *Learning to look: A handbook on classroom observation and teaching model.* Belmont, CA: Wadsworth. Used with permission.

APPENDIX B: Frequencies For Outcomes

Unadjusted frequencies are the raw trend in mean levels of outcomes for each year of implementation before controlling for any other variables. Adjusted frequencies take into account variables that are extraneous to First Things First (e.g., student demographics, cluster affiliation, etc.).

Observed Ratios

	Unadjusted Frequencies			Adjusted Frequencies		
	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL						
Elementary School	16%	19%	12%	17%	20%	14%
Secondary School	7%	13%	11%	12%	27%	18%
RISK						
Elementary School	48%	48%	49%	47%	44%	48%
Secondary School	64%	41%	47%	50%	25%	37%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

Observed Active Learning

	Unadjusted Frequencies			Adjusted Frequencies		
	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL						
Elementary School	35%	30%	29%	31%	24%	26%
Secondary School	30%	27%	42%	32%	25%	40%
RISK						
Elementary School	60%	56%	59%	61%	60%	60%
Secondary School	69%	59%	52%	63%	61%	50%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

Observed Small Grouping

	Unadjusted Frequencies			Adjusted Frequencies		
	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL						
Elementary School	20%	30%	45%	36%	53%	55%
Secondary School	18%	36%	24%	28%	41%	33%
RISK						
Elementary School	69%	38%	33%	52%	19%	24%
Secondary School	75%	48%	35%	55%	32%	24%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

Staff Engagement								
	Unadjusted Frequencies				Adjusted Frequencies			
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Baseline	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL								
Elementary School	20%	24%	26%	27%	26%	37%	42%	32%
Secondary School	16%	23%	28%	24%	25%	42%	47%	33%
RISK								
Elementary School	42%	43%	40%	37%	39%	35%	28%	34%
Secondary School	47%	37%	37%	34%	39%	24%	23%	30%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

Colleague Support								
	Unadjusted Frequencies				Adjusted Frequencies			
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Baseline	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL								
Elementary School	42%	39%	42%	44%	44%	46%	52%	47%
Secondary School	28%	33%	39%	33%	34%	40%	45%	37%
RISK								
Elementary School	35%	40%	38%	35%	37%	39%	35%	35%
Secondary School	52%	51%	42%	42%	48%	42%	40%	41%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

System Support								
	Unadjusted Frequencies				Adjusted Frequencies			
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Baseline	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL								
Elementary School	12%	12%	11%	17%	14%	18%	17%	18%
Secondary School	13%	14%	20%	21%	20%	25%	28%	26%
RISK								
Elementary School	61%	61%	59%	54%	53%	47%	48%	47%
Secondary School	62%	61%	50%	47%	56%	46%	38%	45%

Note: Wyandotte and Washington Clusters Only
Imp indicates implementation

Student Report of Standards								
	Unadjusted Frequencies				Adjusted Frequencies			
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Baseline	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL								
Elementary School	20%	20%	22%	25%	25%	31%	33%	40%
Middle School	6%	9%	10%	11%	6%	7%	7%	9%
High School	3%	3%	4%	5%	4%	3%	4%	8%
RISK								
Elementary School	27%	27%	25%	24%	24%	21%	19%	15%
Middle School	53%	42%	38%	35%	48%	41%	37%	34%
High School	61%	59%	55%	51%	57%	51%	48%	37%

Note: Only Wyandotte and Washington Clusters are included in Imp Year 3
Imp indicates implementation

Student Report of Instruction								
	Unadjusted Frequencies				Adjusted Frequencies			
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Baseline	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL								
Elementary School	22%	22%	22%	22%	22%	22%	23%	23%
Middle School	15%	17%	19%	19%	14%	13%	14%	14%
High School	16%	15%	16%	15%	17%	20%	21%	23%
RISK								
Elementary School	23%	22%	21%	19%	23%	21%	20%	18%
Middle School	40%	33%	29%	30%	38%	35%	31%	31%
High School	39%	42%	37%	41%	40%	41%	37%	37%

Note: Only Wyandotte and Washington Clusters are included in Imp Year 3
Imp indicates implementation

Student Report of Teacher Support										
	Unadjusted Frequencies					Adjusted Frequencies				
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*
OPTIMAL										
Elementary School	22%	25%	29%	36%	32%	27%	36%	38%	48%	49%
Middle School	25%	33%	37%	40%	36%	32%	37%	40%	41%	40%
High School	19%	23%	24%	31%	34%	23%	28%	30%	38%	38%
RISK										
Elementary School	34%	34%	29%	23%	26%	33%	28%	25%	18%	18%
Middle School	32%	24%	19%	20%	19%	27%	22%	18%	19%	18%
High School	31%	27%	23%	19%	17%	28%	24%	21%	17%	18%

***** Only Wyandotte and Washington Clusters are included in Imp Year 4
Imp indicates implementation

Student Engagement										
	Unadjusted Frequencies					Adjusted Frequencies				
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*
OPTIMAL										
Elementary School	33%	33%	34%	33%	32%	34%	32%	37%	37%	39%
Middle School	1%	2%	2%	15%	14%	12%	17%	23%	25%	33%
High School	1%	1%	2%	12%	15%	12%	26%	38%	50%	49%
RISK										
Elementary School	31%	32%	29%	30%	32%	32%	35%	27%	24%	23%
Middle School	42%	33%	29%	23%	21%	37%	33%	25%	27%	22%
High School	47%	41%	36%	27%	22%	42%	44%	34%	31%	26%

***** Only Wyandotte and Washington Clusters are included in Imp Year 4
Imp indicates implementation

Attendance										
	Unadjusted Frequencies					Adjusted Frequencies				
	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*	Baseline	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 4*
OPTIMAL										
Elementary School	62%	61%	63%	61%	63%	63%	63%	64%	63%	66%
Middle School	29%	46%	52%	49%	60%	45%	63%	65%	67%	78%
High School	20%	24%	35%	40%	31%	30%	38%	45%	59%	66%
RISK										
Elementary School	1%	1%	1%	0%	0%	1%	1%	.5%	0%	0%
Middle School	12%	7%	4%	3%	2%	11%	5%	4%	2%	1%
High School	19%	17%	11%	14%	17%	20%	14%	9%	9%	7%

***** Only Wyandotte and Washington Clusters are included in Imp Year 4
Imp indicates implementation

Reading State Test Score						
	Unadjusted Frequencies			Adjusted Frequencies		
	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL						
Elementary School	29%	36%	39%	30%	35%	42%
Middle School	29%	28%	37%	29%	29%	40%
HighSchool	18%	16%	20%	23%	23%	29%
RISK						
Elementary School	39%	34%	31%	42%	36%	27%
Middle School	35%	32%	29%	42%	38%	30%
High School	46%	53%	51%	57%	60%	56%

Note: Implementation year 1: Harmon and Schlagle only; Imp year 2: Harmon, Schlagle, Washington; Imp year 3: All four clusters
Imp indicates implementation

Math State Test Score						
	Unadjusted Frequencies			Adjusted Frequencies		
	Imp Year 1	Imp Year 2	Imp Year 3	Imp Year 1	Imp Year 2	Imp Year 3
OPTIMAL						
Elementary School	33%	37%	39%	29%	33%	38%
Middle School	12%	23%	18%	18%	27%	25%
HighSchool	5%	5%	7%	4%	4%	6%
RISK						
Elementary School	35%	34%	32%	44%	41%	34%
Middle School	63%	51%	57%	64%	56%	57%
High School	74%	67%	68%	75%	69%	69%

Note: Implementation year 1: Harmon and Schlagle only; Imp year 2: Harmon, Schlagle, Washington; Imp year 3: All four clusters
Imp indicates implementation

Graduation and Dropout						
	Unadjusted Frequencies			Adjusted Frequencies		
	Graduating Class 2001	Graduating Class 2002	Graduating Class 2003	Graduating Class 2001	Graduating Class 2002	Graduating Class 2003
Graduation Rate						
Wyandotte	70%	81%	82%	79%	87%	87%
Washington	67%	75%	80%	72%	80%	84%
Harmon	67%	79%	84%	74%	84%	86%
Schlagle	64%	75%	81%	66%	77%	80%
Dropout Rate						
Wyandotte	19%	16%	6%	24%	9%	2%
Washington	8%	7%	6%	22%	4%	2%
Harmon	32%	17%	11%	27%	11%	5%
Schlagle	12%	7%	6%	27%	19%	13%

APPENDIX C: Relative Change In Outcomes After Three Years Of Implementation

Table C-1: Relative Improvement in Structural Implementation Outcomes Wyandotte and Washington Clusters

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Observed Ratios (Year 1 to Year 3)	*.76*	ns	ns	.70
SLC Stability (Year 2 to Year 3)	1.99	.04	4.02	.045

Values shown in *italics* are not in the expected direction
ns ***indicates not significant***

Ratios:
Optimal = LE 15:1
Risk = GE 21:1 elementary; 23:1 secondary

Stability:
Optimal = students in SLC 2 or more years
Risk = students in SLC fewer than 2 years

Table C-2: Relative Improvement from Baseline to Year 3 of Implementation in Instructional Implementation Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Observation of Small Grouping (Year 1-Year 3 WY/WA)	2.22	.48	1.28	.47
Observation of Active Learning (Year 1-Year 3 WY/WA)	ns	ns	1.43	.76
Student Report of Instruction (Baseline to Year 3, all clusters)	1.01	.77	H=1.37	M=.80 H=.93
Student Report of Standards (Baseline to Year 3, all clusters)	1.79	.61	M=1.50 H=2.36	M=.71 H=.63

H ***indicates High School***
M ***indicates Middle School***
ns ***indicates not significant***

Observation of Small Grouping:
Optimal = students spend two-thirds or more of observed time in small groups
Risk = students spend one-third or less of observed time in small groups

Observation of Active Learning:
Optimal = students spend 60 percent or more of observed time in active learning
Risk = students spend less than 25 percent of observed time in active learning

Student Report of Instruction:
Optimal = at least 9 of 12 items answered "Sort of True" or "Very True"
Risk = at least 4 of 12 items answered "Not Very True" or "Not At All True"

Student Report of Standards:
Optimal = at least 7 of 10 items answered "Most of the Time" or "Almost Always"
Risk = at least 6 of 10 items answered "Not Very Often" or "Almost Never"

Table C-3: Relative Improvement from Baseline to Implementation Year 3 in Student Intermediate Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Student Report of Teacher Support	1.78	.57	M=1.62 H=1.83	M=.62 H=.55
Student Engagement	1.08	.88	M=14.87 H=34.24	M=.63 H=.61

H ***indicates High School***
M ***indicates Middle School***

Teacher Support:
Optimal = mean of 3.75 or higher on 8 items for elementary students; 3.25 or higher for secondary
Risk = Mean of 3.0 or lower on 8 items for elementary students; 2.5 or lower for secondary

Engagement:
Optimal = mean of 3.75 or higher on 10 items for elementary students; 8 items for secondary
Risk = mean of 3.25 or lower on 10 items for elementary; mean of 3.0 or lower on 8 items for secondary

Table C-4: Relative Improvement from Baseline to Year 3 of Implementation in Staff Intermediate Outcomes in the Wyandotte and Washington Clusters

	ELEMENTARY		SECONDARY	
	Increased Optimal	Decreased Risk	Increased Optimal	Decreased Risk
Colleague Support	ns	ns	ns	.84
System Support	1.42	.89	1.45	.79
Staff Engagement	1.36	.86	1.51	.74

ns ***indicates not significant***

Colleague Support:
Optimal = mean of 3.5 or higher on 6 items
Risk = mean of 3.0 or lower on 6 items

System Support:
Optimal = mean of 3.5 or higher on 10 items
Risk = mean of 3.0 or lower on 10 items

Staff Engagement:
Optimal = mean of 3.5 or higher on 12 items
Risk = mean of 3.0 or lower on 12 items

Table C-5: Relative Improvement in Achievement Outcomes

	ELEMENTARY		SECONDARY	
	Increased Optimal	**Decreased Risk**	**Increased Optimal**	**Decreased Risk**
Attendance	*.98*	.18	M=1.87 H=2.53	M=.21 H=.41
State Reading Test (SY00-01 to 02-03)	1.75	.43	M=1.97 H=1.61	M=.31 H=.78
State Math Test (SY00-01 to 02-03)	1.66	.51	M=2.12 H=1.67	M=.71 H=.88
High School Graduation Rate (Graduating Class 2001-Graduating Class 2003)	NA	NA	Wy=1.05 Wa=1.10 Ha=1.06 Sc=1.03	NA
High School Dropout Rate	NA	NA	NA	Wy=.08 Wa=.03 Ha=.07 Sc=.07

Values in *italics* indicate significant relationship not in the expected direction
H ***indicates High School***
M ***indicates Middle School***
NA ***indicates not applicable***

Attendance:
Optimal = 1 or fewer absences per month
Risk = 1 or more absences per week

State Tests:
Optimal = Proficient or above
Risk = Unsatisfactory

Graduation:
Optimal = Graduates within 5 years

Youth Development Strategies, Inc.

429 Fulton Street
Philadelphia, PA 19147
www.ydsi.org