

Urban School Reform: Year 4 Outcomes for the
Knowledge is Power Program in an Urban Middle School

Aaron J. McDonald, Steven M. Ross, Jane Abney, and Todd Zoblotsky

Center for Research in Educational Policy

The University of Memphis

Paper presented at the annual meeting of the
American Educational Research Association in New York, NY March 2008

Abstract

The present longitudinal study examined the fourth year effects of the Knowledge is Power Program, which is specifically designed to raise academic achievement of at-risk urban middle school students via interventions such as extended learning time, rigorous curriculum, after-school access to teachers, and extensive family-school connection. This study individually matched 165 KIPP students with control students of the same ethnicity, SES, gender, and ability who attended demographically similar schools. While the majority of the comparisons directionally favored KIPP students this year, only the 5th grade mathematics comparison was significant. The loss of the school principal and increased student behavior issues are interpreted as likely contributing factors.

Introduction

The No Child Left Behind Act of 2002 reflected an ongoing history of federal reform programs in the United States that have emphasized improving the achievement of at-risk students. A large portion of these programs have been aimed at assisting urban districts, which serve a large portion of the nation's disadvantaged students (Snipes & Casserly, 2004; U.S. Department of Education, 1999). This targeted assistance makes sense because the largest 100 districts in the nation contain 17% of all public schools, employ 21% of all public school teachers, serve 30% of the nation's disadvantaged students, and 40% of its minority students (Snipes & Casserly, 2004).

Despite continued funding for programs aimed at assisting disadvantaged students, "closing the achievement gap" between disadvantaged and non-disadvantaged students remains a priority. While there certainly have been successes in urban reform, many reform initiatives have been inhibited because of factors such as low teacher buy-in, inadequate resources, insufficient professional development, conflicting district policies, and failure to demonstrably raise student achievement (Berends, Kirby, Naftel, & McKelvey, 2002; Ross, 2001; Snipes & Casserly, 2004). A recent example is noted by Ross (2003) who analyzed the failure of urban reforms in Memphis and Toledo. Lacking at many schools was a clearly stated mission, a safe environment, high expectations, instructional leadership, monitoring of progress, and effective communications.

For a variety of reasons, implementation of school reforms has been demonstrated to take place more slowly in high schools and middle schools in comparison with elementary schools (Bodilly, 1998; Bodilly & Berends, 1999). In the higher grades, curricula tend to be more departmentalized and the teachers more resistant to experimenting with innovative programs and

instructional strategies (Smith et al., 1997). The middle grades also bring challenges associated with the psychological, social, and biological changes that impact students during adolescence (Simmons & Blyth, 1987). Ability to cope with stress and emotional stability in general tend to vary in grades 5-7 (Larson, Moneta, Richards, & Wilson, 2002). Not surprisingly, the need for supportive middle school environments has been strongly emphasized in research studies (see Roeser & Eccles, 1998; Way & Robinson, 2003).

In an effort to create a reform program geared more closely to the needs of at-risk urban students, two student participants in Teach for America developed the Knowledge Is Power Program (KIPP) in 1994 (KIPP, n.d.). KIPP is grounded on five foundational “pillars”: High Expectations, Choice and Commitment by Families, More Time to Learn, Power to Lead, and Focus on Results. The KIPP philosophy is operationalized through an emphasis on (a) graduating from both high school and college; (b) spending more time in the classroom through extended day, extended year, and Saturday classes; (c) assignment of 2-3 hours of homework each night; (d) after-school access to teachers via cell phones; and (e) extensive professional development for school leaders.

The present study used a mixed-method design to examine fourth-year outcomes from the implementation of KIPP at an inner-city middle school. In this approach, quantitative and qualitative techniques are both used in the same study (Onwuegbuzie & Teddlie, 2003). The rationale underlying this method is to offer a logical and pragmatic evaluation design that incorporates both induction and deduction (Johnson & Onwuegbuzie, 2004). The combination of data sources is intended to build on strengths of each type of data, while minimizing weaknesses of any single approach. Similarly, this approach can strengthen the validity of results through triangulation (Frechtling & Sharpe, 1997). The multiple data sources employed

in the study focused on school climate, perceptions by key stakeholders, systematic monitoring of program implementation progress, classroom observations, and student achievement results on state-mandated assessments.

Method

Design and Participants

Both descriptive and inferential analyses were conducted, the latter involving a quasi-experimental comparison of KIPP and matched comparison groups on student achievement. Multiple data sources for the present study, described in greater detail below, included (a) the School Observation Measure; (b) the School Climate Inventory; (c) the Comprehensive School Reform Teacher Questionnaire; (d) teacher focus groups; (e) principal interview; (f) implementation benchmark review; (g) the KIPP Parent Questionnaire; (h) student focus group; and (i) student-level test scores (Language Arts/Reading and Math) on the Tennessee Comprehensive Assessment Program Achievement Test (TCAP/AT).

Participants in the study included KIPP parents, faculty and staff members, as well as both new and continuing KIPP students. KIPP students were comprised of the following cohorts: (a) fifth ($n = 42$), sixth ($n = 19$), seventh ($n = 12$), and eighth graders ($n = 9$) who had completed their first year at the school; (b) sixth ($n = 28$), seventh ($n = 7$), and eighth graders ($n = 3$) who had completed their second school year (e.g., students who enrolled at KIPP in the 2004-05 academic year); (c) seventh ($n = 26$) and eighth ($n = 3$) graders who had completed their third year (e.g., students who enrolled at KIPP in the 2003-04 academic year); and (d) eighth graders ($n = 16$) who had completed their fourth year (e.g., students who enrolled as fifth graders during the 2002-03 academic year). All (100%) were African-American, and approximately 92% were eligible for free or reduced price lunch. For the achievement analysis, a rigorous matched

treatment-comparison group design was employed by establishing 165 student-level matched-pairs. The potential comparison group pool was selected from students at schools that were highly comparable to KIPP and to each other in both student and school demographics. Each KIPP student was matched with a comparison student based on criteria including: grade level, race, gender, socio-economic status (SES), and pretest scores in Reading/Language Arts and Mathematics.

Instrumentation

School Observation Measure (SOM). The SOM was developed to determine the extent to which different common and alternative teaching practices are used throughout an entire school (Ross, Smith, & Alberg, 1998). The procedure involves observers' visiting 10-12 randomly selected classrooms, for 15 minutes each, during a three-hour visitation period. The observer examines classroom events and activities descriptively, not judgmentally. Notes are taken relative to the use or nonuse of 24 target strategies. At the conclusion of the three-hour visit, the observer summarizes the frequency with which each of the strategies was observed across all classes in general on a data summary form. The frequency is recorded via a 5-point rubric that ranges from (0) Not Observed to (4) Extensively. Two global items are used to rate, respectively, the level of academically-focused instructional time and degree of student attention and interest.

The SOM strategies include traditional practices (e.g., direct instruction and independent seatwork) and alternative, predominantly student-centered methods associated with educational reforms (e.g., cooperative learning, project-based learning, inquiry, discussion, using technology as a learning tool). The strategies were originally identified through surveys and discussions involving policymakers, researchers, administrators, and teachers, as those most useful in

providing indicators of schools' instructional philosophies and implementations of commonly used reform designs (Ross, Smith, Alberg, & Lowther, 2004).

To ensure the reliability of data, observers receive one full day of training, a manual providing definitions of terms, examples and explanations of the strategies, and a description of procedures for completing the instrument. In a reliability study (Lewis, Ross, & Alberg, 1999), pairs of trained observers selected the identical overall response on the five-category rubric on 67% of the items and were within one category on 95% of the items. In a reliability study using Generalizability Theory, Sterbinsky and Ross (2003) found reliability at the .74 level for 5 SOMs conducted at individual schools. Reliability increased to .82 with 8 SOMs and to .85 with 10 SOMs conducted at a school.

School Climate Inventory. The School Climate Inventory (SCI) consists of seven dimensions logically and empirically linked with factors associated with effective school organizational climates (Butler & Alberg, 1989). Each scale contains seven items, with 49 statements comprising the inventory. Responses are scored through use of Likert-type ratings [strong disagreement (1) to strong agreement (5)]. Scale means can range from 1 to 5 with higher scores being more positive. Additional items solicit demographic information.

Face validity of the school climate items and logical ordering of the items by scales were established by the research team during the development of the inventory (Butler & Alberg, 1989). Subsequent analysis of responses collected through administration of the inventory in a variety of school sites substantiates validity of the items. Dimension descriptions and current internal reliability coefficients on the seven dimensions of the inventory, obtained using Cronbach's alpha, are as follows: Order – the extent to which the environment is ordered and appropriate student behaviors are present ($\alpha = .84$); Leadership – the extent to which the

administration provides instructional leadership ($\alpha = .83$); Environment – the extent to which positive learning environments exist ($\alpha = .81$); Involvement – the extent to which parents and the community are involved in the school ($\alpha = .76$); Instruction – the extent to which the instructional program is well developed and implemented ($\alpha = .75$); Expectations – the extent to which students are expected to learn and be responsible ($\alpha = .73$); and Collaboration – the extent to which the administration, faculty, and students cooperate and participate in problem solving ($\alpha = .74$).

Comprehensive School Reform Teacher Questionnaire. The Comprehensive School Reform Teacher Questionnaire (CSRTQ) was designed to assess teachers' perceptions of school conditions that affect capacity to enact reforms (Ross & Alberg, 1999). The CSRTQ contains 28 items to which teachers respond using a five-point Likert-type scale ranging from strongly disagree (1) to strongly agree (5). In a second section, respondents report their perceived progress toward implementation benchmark goals. In a third section, they provide open-ended comments regarding the positive and negative aspects of their school's reform program. Face validity and logical ordering of the items by scales were established by the research team during the development of the inventory (Ross et al., 1997). Principal component analyses of responses collected through administration of the inventories in multiple restructuring and control sites substantiated the instrument's five scales. Dimension descriptions and current internal reliability coefficients on the five scales, obtained using Cronbach's alpha, are as follows: Support - the extent to which the school receives effective professional development and support to implement its CSR program ($\alpha = .82$); Capacity/Resources - the extent to which planning time, materials, technology, and faculty are available at the school ($\alpha = .70$); Pedagogy - the extent to which classroom practices, materials, and technology use have changed at the school ($\alpha = .75$);

Outcomes - the extent to which positive student, faculty, and parent/community outcomes have occurred as a result of CSR ($\alpha = .90$); Focus - the extent to which elements of the school's educational program are integrated, evaluated, and supported by school stakeholders ($\alpha = .83$). More recently, Nunnery, Ross, and Sterbinsky (2003) conducted a construct validation study using the CSRTQ. This study showed the *support* and *capacity* dimensions to account for 84% of the variance in *focus*, which in turn had strong direct effects on both *pedagogy* and *outcomes*. Importantly, 84% of the variance in *outcomes* was explained by *focus* and *pedagogy* combined.

KIPP Parent Questionnaire. The Parent Questionnaire is a sixteen-item survey developed to gather parent perceptions concerning the KIPP Academy (Ross & McDonald, 2003). The first twelve questions examine parent and student adjustment to KIPP, level of comfort and satisfaction, extent of parental involvement, and interaction with teachers. Parents responded to these questions using a five point Likert-type scale ranging from strongly disagree (1) to strongly agree (5). The next four questions are "open-ended" and provide parents with an opportunity to make comments concerning strengths and weaknesses of the program, parental communication, and impact on students. The instrument was evaluated for content and face validity based on reactions by a diverse review panel consisting of principals, teachers, and researchers.

Focus groups and interviews. To supplement the survey data, we conducted a principal interview, a teacher focus group, and three student focus groups including 15 different students. The time period for each was approximately one hour. Teacher and student participants were randomly selected to participate. A semi-structured protocol, involving standard questions with flexibility for follow-up on selected responses, was used. In all three protocols, the basic questions concerned experiences during the year, differences from other schools, reactions to

major school components (e.g., teaching methods, extended-day, curriculum, parent involvement), perceived strengths and weaknesses, and recommendations for improvement.

Program Implementation Benchmarking. A highly critical component of formative evaluation and improvement planning is the development of individualized “Implementation Benchmarks.” As part of the benchmark development process, the evaluators and the school staff develop statements or goals for each major program component. Each statement is then accompanied by a specific *indicator* and *evidence* for implementation Phases I (Beginning), II (intermediate), and III (Full). The typical process is to complete development of the implementation benchmarks, refine them if needed during the year, and then, by spring, engage the school staff in evaluating progress and specifying program goals for the following year. Based on those goals, the benchmarks are continually revised and refined. Consequently, participating schools are continually aware of all program components, implementation progress, directions for school improvement directed by data, and shared faculty-administrator decision-making. In general, the benchmarking process accomplishes the following: Documents the primary operational components of a school program to increase understanding of both the overall program and individual school goals relative to implementation rate and scope; Engages the entire school staff in discovering, developing, reflecting on, and refining the school’s programs; Provides a framework for evaluating progress from beginning to full implementation; and Provides a tool for communicating implementation status and progress.

Tennessee Comprehensive Assessment Program Achievement Test (TCAP/AT). The TCAP/AT, administered in grades 3-8, is the state-mandated achievement assessment in Tennessee. Prior to 2003, TCAP/AT consisted exclusively of a norm-referenced test (NRT) consisting of subtests in Reading, Language Arts, Mathematics, Science, and Social Studies.

Starting in spring, 2003 and continuing in spring, 2004, the NRT was supplemented by a criterion-referenced test (CRT) consisting of subtests in Reading/Language and Mathematics. In 2005, only the CRT portion was administered.

The NRT portion of TCAP/AT consists of the TerraNova or CTBS-5 (CTB/MacMillan/McGraw Hill, 1997) for five subjects (Language Arts, Reading, Mathematics, Science, and Social Studies). The purpose of the CRT portion of TCAP/AT is to measure student performance according to specific subject area standards rather than to the performance of other test takers. Accordingly, the TCAP/AT-CRT items are directly aligned with Tennessee's Content Standards and Performance Indicators. Similar to the NRT, a multiple-choice format is used. Student performance is assessed relative to categories of "Below Proficient," "Proficient" and "Advanced" based on scale score cutoffs.

Procedure and Analyses

During the school year, researchers visited the KIPP Academy six times to conduct half-day classroom observations. In the spring, the evaluation team administered surveys and/or interviews to the faculty, students, and parents. Quantitative data were analyzed via descriptive summaries of frequency counts and means, and in the case of student achievement, via multivariate analysis of covariance (MANCOVA) comparing KIPP and comparison group scores on TCAP measures. Qualitative analyses, guided by Miles and Huberman's (1994) analysis model, were performed on open-ended survey and interview responses. The procedure consisted of transcribing ideas and concepts, deriving patterns and categories, identifying themes, and revision and refinement based on member checking and interrater review.

Results

The results are presented in three sections which focus on (1) perceptual data from principals, teachers, parents, and students, (2) observations of teaching, and (3) student achievement.

Participant Reactions

School Climate. Climate has declined over the past four years, but remains positive overall (see Table 1). Six of the seven dimensions had means near or over 4.0 (based on a 1 – 5 scale). The Order dimension received the lowest mean (2.98) and also showed the greatest decline over the four years. The climate averages for KIPP were generally on par with averages for other secondary schools (no statistically significant differences). One respondent noted that “[the school] has had an unstable administration” which has impacted climate.

Table 1.
SCI-R means during the last four academic years and national average (norm) for secondary schools

Dimension	2002 - 2003	2003 - 2004	2004 - 2005	2005 - 2006	National Norm
Collaboration	4.37	4.22	4.16	4.04	3.71
Environment	4.51	4.10	4.26	3.95	3.73
Expectations	4.66	4.40	4.49	4.08	3.82
Instruction	4.71	4.32	4.25	3.98	4.06
Involvement	4.54	4.29	4.21	4.00	3.63
Leadership	4.63	4.30	4.16	3.88	3.94
Order	4.11	3.87	3.25	2.98	3.26
OVERALL	4.50	4.21	4.11	3.84	3.73
Number of Respondents	5	9	11	7	

Scale: 1 = Strongly Disagree and 5 = Strongly Agree

Teacher Questionnaire. There was a general decrease in the five dimension means across years (see Table 2). In terms of the dimensions measured on the CSRTQ, teachers were most positive regarding Support (M = 3.66) and Outcomes (M = 3.57). Teachers were less positive regarding Pedagogy (M = 3.40) and Capacity/Resources (M = 3.39). Despite the Capacity/Resources dimension being lower than the other dimensions in 2005-2006, this dimension actually increased over the previous year more than the other dimensions did. Within Capacity/Resources, more than half of the respondents felt that the school had sufficient faculty

and staff to fully implement the CSR program (57.1% agreement) and, nearly three-fourths (71.4%) felt materials needed to implement the program are readily available. Overall, KIPP teacher perceptions were generally comparable to national norms for the CSRTQ.

The primary areas of concern in the open-ended comments involved parents' negative attitudes, no consequences for poor student behavior, and the lack of strong leadership for the past several years. On a positive note, teachers were pleased with a number of aspects of the school's program such as, the learning strategies, the idea (or potential) of strict discipline, the open communication in general, the teamwork (team/family atmosphere), the college preparatory plan, and the longer hours coupled with rigorous instruction.

Table 2.
CSRTQ Dimension means during the last four academic years and national average (norm) for secondary schools

Dimension	2002 - 2003	2003 - 2004	2004 - 2005	2005 - 2006	National Norm
Capacity/Resources	4.19	3.45	3.14	3.39	3.38
Focus	4.65	4.03	3.47	3.51	3.54
Outcomes	4.86	4.07	3.72	3.57	3.26
Pedagogy	4.10	3.93	3.62	3.40	3.26
Support	4.55	3.70	3.71	3.66	3.59
OVERALL	4.47	3.84	3.53	3.51	3.41
Number of Respondents	4	6	11	7	

Scale: 1 = Strongly Disagree and 5 = Strongly Agree

Parent Questionnaire. Table 3 presents the percentage of agreement with questionnaire items (longitudinal findings are not presented in this table because the questionnaire was revised; the survey results for years 1-3 can be found in the Appendix). Overall, parents were very positive regarding the KIPP Academy. When the parents were asked if they were pleased that their child attended the school and if they felt that KDA had high expectations for student achievement, 100% of the respondents agreed or strongly agreed. Additionally, three questions regarding the academics at KDA also received praise according to the questionnaire: 94.4% of parents felt that the teachers provided high-quality instruction, the curriculum was challenging, and the curriculum was appropriate to their child's needs.

The two items with lower levels of agreement (72.2% agreement) concerned the aesthetics of the school: the school grounds are well maintained and the classrooms and hallways are attractive. Parents also had some reservations about the principal being a good leader, another item that only garnered 72.2% agreement. This response may have been influenced by the school's having an interim principal in place for the last four months of the academic year. Transportation service was another area about which parents agreed less, with only 61.1% indicating they thought they were adequate.

In the open-ended comments section of the questionnaire, parents reiterated that they believed some strengths of the school were the high-quality instruction, the involvement and dedication of the teachers, and the high expectations placed on the students. When asked what needs to be improved, the majority of the parents cited poor student conduct and lack of strong student discipline as the areas most in need of change. Overwhelmingly, parents who responded to the KPQ-R seemed to be quite pleased with the Academy. Many concluded their survey with comments such as, “[the] Academy is an excellent school,” “the best thing to happen to the public school system,” “[this] is a great school,” and “keep up the great work!”

Table 3. *Parent Questionnaire Data Summary*
Year 4 (2005-2006) N = 18

KIPP Parent Questionnaire - Revised (KPQ-R)	% Strongly Agree and Agree	% Neutral	% Strongly Disagree and Disagree
Teachers provide high-quality instruction.	94.4	5.6	0.0
The curriculum is challenging.	94.4	5.6	0.0
The curriculum is appropriate to my child's needs.	94.4	5.6	0.0
The principal is a good leader.	72.2	22.2	5.6
Rules for student behavior are consistently enforced at this school.	77.8	5.6	11.1
Classrooms and hallways are attractive.	72.2	22.2	5.6
The school grounds are well maintained.	72.2	27.8	0.0
This is a safe school.	88.9	11.1	0.0

KIPP Parent Questionnaire - Revised (KPO-R)	% Strongly Agree and Agree	% Neutral	% Strongly Disagree and Disagree
There are high expectations for student achievement.	100.0	0.0	0.0
Parents regularly receive information about their child's progress (test results, conferences, etc.).	94.4	5.6	0.0
School personnel treat parents with respect.	88.9	11.1	0.0
Parent involvement is encouraged at this school.	94.4	5.6	0.0
Teachers are available if parents need to talk to them.	94.4	5.6	0.0
Transportation services are adequate.	61.1	27.8	5.6
My child likes attending this school.	88.9	11.1	0.0
I am pleased that my child is attending this school.	100.0	0.0	0.0

Focus Groups, Interview, and Benchmark Review Sessions. Similar to the questionnaires, areas across respondent groups that were viewed as strengths of the school were: (a) high expectations for children, (b) extended learning time, (c) impact on student learning, and (d) commitment to school goals by all stakeholders. During this academic year, all stakeholders agreed with and supported the school's goals of preparing "at risk" students for college, encouraging students to give back to the community, teaching students to be good problem solvers, and helping students to be productive members of society. Some of the strategies used to reach these goals include: incorporating the idea of college into the school culture by constantly talking about it and by academically preparing the students for college, communicating with the students to try to help them work through their problems and look for solutions, helping to develop foundational skills that students will use in high school and college, and sponsoring trips that provided diverse experiences to students who might not have been given such opportunities otherwise. Along with this, all stakeholders felt the extended amount of time spent on task afforded students a higher rate of retention and recall, therefore building better academic mastery.

Also similar to last year, there were clear themes that were seen as barriers to success this year. These were (a) student misbehavior and (b) changes in the school principalship. Student misbehavior was clearly noted as being problematic across respondent groups. Teachers noted that much time was spent on classroom management issues. Teachers also noted that the discipline policy was not clearly stated or enforced. Encouragingly, students reported fewer physical altercations in comparison with last year, but did cite inconsistent enforcement of school rules as well as perceived teacher “favoritism” toward certain students.

In addition to student behavior problems, another issue has been instability in school leadership. Because of chronic health problems, the beloved founding principal had to step down from the principalship. This resulted in a series of interim principals from KIPP national. This leadership discontinuity reportedly had a major impact on morale, direction, and cohesiveness amongst the faculty and staff. When a new principal was finally hired in the spring of 2005, she implemented many curricular and organizational changes. The impact of the principal was summed up by one teacher who commented “the new administrator has made many positive changes and this school feels more like a KIPP school everyday” and “it is a blessing to be here!” Unfortunately, this principal had to leave the school in February 2006 due to medical reasons, which again resulted in an interim principal. This change negatively impacted morale and the forward momentum that was seen at the end of the 2004-05 academic year. Teacher focus group respondents stated that they did not feel that they had one person or a core leadership group that was giving everyone the same message; instead, lots of voices were coming at them, voices that were rarely “in sync” with one another. One teacher stated that training and support provided for the program this academic year was sporadic. The teachers also felt that the inconsistency and lack of stability in the administration crippled their ability to

excel in their job. They stated that there was no feedback on how or what they were doing and that there was not enough training and support this year for them to be as effective as they wanted to be. The teachers also noted that the lack of strong leadership impacted follow-through on school goals. Whether on the part of the teachers, students, parents, or the administration, the group agreed that it does not seem like they ever get to the finish line. The overwhelming feeling was that it all just got “lost in the middle” somewhere. Nonetheless, the teachers stated that they remained committed to the school’s goals and truly believed that they can have a positive impact on students.

Classroom Observations

Table 4 presents the means for each strategy included on the SOM. In Year 4, KIPP teachers tended to employ traditional instructional strategies such as direct instruction and independent seatwork. Teachers supplemented these practices with items including independent inquiry/research on the part of students. Additionally, the teachers also transitioned to a coach/facilitator role during lessons. There was some evidence of student centered strategies such as cooperative/collaborative learning and student discussion, but these strategies were not utilized extensively. Technology was observed rarely or not at all during the visits. The level of academically focused class time was high during the majority of visits. High levels of student attention, interest, and engagement were observed frequently during the majority of visits. The teaching practices employed were closely comparable to the national norms, although KIPP had higher parent/community involvement in learning activities, a higher level of direct instruction and sustained reading, and more inquiry/research on the part of students.

Table 4. Comparison of KIPP SOM Item Means to National Norms

Scale: (0 = Not Observed to 4 = Extensively)

Dimension	2003 (n=6)		2004 (n=6)		2005 (n=6)		2006 (n=6)		Secondary Norm (n=370)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Instructional Orientation</i>										
Direct instruction	2.67	1.21	1.83	0.75	3.17	0.75	3.17	0.75	2.89	1.06
Team teaching	0.00	0.00	0.17	0.41	0.00	0.00	0.33	0.52	0.42	0.61
Cooperative/collaborative learning	1.67	0.82	2.17	1.60	1.17	0.75	0.67	0.82	1.04	1.14
Individual tutoring	0.33	0.82	0.33	0.52	0.33	0.52	0.17	0.41	0.43	1.16
<i>Classroom Organization</i>										
Ability groups	0.17	0.41	0.67	0.52	0.67	0.52	0.00	0.00	1.18	1.35
Multi-age grouping	0.00	0.00	0.67	0.52	0.33	0.52	0.00	0.00	1.21	1.52
Work centers	0.50	0.84	1.67	3.61	0.33	0.52	0.00	0.00	0.41	0.77
<i>Instructional Strategies</i>										
Higher-level instructional feedback	1.17	1.33	1.17	1.17	2.00	1.10	0.67	0.52	1.55	1.22
Integration of subject areas	0.50	0.84	0.00	0.00	0.67	1.21	0.00	0.00	0.54	0.97
Project-based learning	1.17	1.33	2.17	1.83	0.50	0.55	0.00	0.00	0.71	1.02
Use of higher-level questioning strategies	1.50	1.05	1.00	0.89	2.00	1.10	0.67	0.52	1.55	1.11
Teacher acting as a coach/facilitator	2.17	0.98	2.33	1.03	2.17	0.98	2.50	1.05	2.54	1.18
Parent/community involvement in learning activities	0.17	0.41	0.17	0.41	0.17	0.41	0.17	0.41	0.11	0.56
<i>Student Activities</i>										
Independent seatwork	1.83	0.98	2.33	1.03	3.83	0.41	2.33	0.82	2.76	1.01
Experiential, hands-on learning	0.33	0.52	0.67	0.52	1.00	0.89	0.83	0.41	0.85	1.06
Systematic individual instruction	0.00	0.00	0.33	0.52	0.67	0.82	0.00	0.00	0.32	0.82
Sustained writing/composition	1.00	0.89	0.17	0.41	0.33	0.52	0.33	0.52	0.67	1.07
Sustained reading	0.17	0.41	0.00	0.00	0.50	1.22	0.83	0.41	0.74	1.28
Independent inquiry/research on the part of students	0.17	0.41	0.00	0.00	0.33	0.52	0.83	0.82	0.46	0.91
Student discussion	0.33	0.52	0.67	1.21	0.83	0.75	0.67	0.82	0.71	1.30
<i>Technology Use</i>										
Computer for instructional delivery	0.33	0.82	0.00	0.00	0.17	0.41	0.33	0.52	0.65	0.85
Technology as a learning tool or resource	0.33	0.52	2.00	1.67	0.50	0.55	0.00	0.00	0.78	0.94
<i>Assessment</i>										
Performance assessment strategies	0.17	0.41	0.00	0.00	0.17	0.41	0.00	0.00	0.69	1.26
Student self-assessment	0.17	0.41	0.00	0.00	0.33	0.82	0.00	0.00	0.45	1.33
<i>Summary Items</i>										
High academically focused class time	3.17	0.41	2.83	0.98	3.00	0.63	3.00	0.00	2.94	0.86
High level of student attention/interest/engagement	3.83	0.41	2.83	0.98	2.50	0.84	2.83	0.41	2.65	0.89

Student Achievement

As previously described, a quasi-experimental design with matched treatment and comparison groups was employed. The outcome measures for the present study were the Reading/Language Arts and Mathematics subtests on TCAP/AT. Two types of data were analyzed. For the primary analyses, we used the number of items answered correctly on each subtest. These data represented an “interval” measure that is more precise than simply knowing the students’ performance levels (e.g., Below Proficient, Proficient, etc.), as established by state raw score cut-offs. However, because the latter indices are those publicly reported and used to determine schools’ Adequate Yearly Progress (AYP) under the No Child.

Procedure and Analyses

One-way Analysis of Variance (ANOVA) was conducted on baseline achievement data for Reading/LA and Math within each grade level to ensure comparability of KIPP and control group cohorts. Correlations were also performed to test the relationship between pre- and post-implementation achievement scores. Multivariate Analysis of Covariance (MANCOVA) was used to assess the impact of KIPP membership on student Math and Reading/LA achievement scores, with students’ baseline (pre-charter school) achievement score used as the covariate. Finally, chi-square analyses were conducted to compare proficiency levels attained by KIPP and control students for each grade by subject area and cohort. The first analysis evaluated three performance levels (Below Proficient, Proficient, and Advanced).

While looking at differences in these three proficiency levels can show more detailed changes in student performance, it is also meaningful to consider whether there are increases in the overall proportion of students meeting basic proficiency. For example, a school may not have made statistically significant changes in the number/percent of students moving from one

specific level to the other (e.g., Proficient to Advanced), but may have made significant changes in the number/percent moving from not being proficient to proficient, which is an important accountability issue under No Child Left Behind. As a result, a second chi square analysis collapsed the three performance levels into two for analysis: Pass (Proficient and Advanced) and Fail (Below Proficient).

Achievement Results

As shown in Table 5, one-way ANOVA confirmed the adequacy of the matching process in each subject area by cohort and grade. Pre- and post-implementation means and standard deviations are summarized in Tables 5 and 6, respectively. Effect sizes were also obtained for (1) each grade combined and subject area at pre-implementation and (2) for each grade combined, cohort, and subject area at the post-implementation. Cohen's *d* effect size was computed as the mean difference (treatment – control) divided by the pooled standard deviation.

Table 5. Baseline Means, Standard Deviations, and Effect Sizes

Grade and Cohort	Group	Reading/LA			Math		
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
5 th Grade (1 st Year)	KIPP students	42	33.000	8.462	42	35.857	7.719
	Control students	42	32.500	8.034	42	35.024	7.003
	Effect Size	<i>d</i> = 0.06			<i>d</i> = 0.11		
	One-way ANOVA	<i>F</i> = 0.08; <i>p</i> = 0.7819			<i>F</i> = 0.27; <i>p</i> = 0.6057		
6 th Grade (1 st Year)	KIPP students	19	46.158	13.849	19	43.579	16.648
	Control students	19	46.053	12.878	19	43.421	17.056
6 th Grade (2 nd Year)	KIPP students	28	41.000	12.083	28	38.536	14.263
	Control students	28	40.571	11.432	28	38.571	13.836
6 th Grade (Combined)	KIPP students	47	43.085	12.934	47	40.574	15.301
	Control students	47	42.787	12.206	47	40.532	15.231
	Effect Size	<i>d</i> = 0.02			<i>d</i> = 0.00		
	One-way ANOVA	<i>F</i> = 0.01; <i>p</i> = 0.9088			<i>F</i> = 0.00; <i>p</i> = 0.9892		

Table 5, continued.

Grade and Cohort	Group	Reading/LA			Math		
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
7 th Grade (1 st Year)	KIPP students	12	43.667	10.773	12	37.833	12.401
	Control students	12	43.167	9.408	12	38.250	11.169
7 th Grade (2 nd Year)	KIPP students	7	35.857	21.287	7	38.857	13.196
	Control students	7	36.429	20.598	7	38.857	12.602
7 th Grade (3 rd Year)	KIPP students	26	41.885	12.701	26	39.538	11.884
	Control students	26	42.154	13.129	26	39.962	10.716
7 th Grade (Combined)	KIPP students	45	41.422	13.744	45	38.978	11.957
	Control students	45	41.533	13.529	45	39.333	10.894
	Effect Size	<i>d</i> = - 0.01			<i>d</i> = - 0.03		
	One-way ANOVA	<i>F</i> = 0.00; <i>p</i> = 0.9693			<i>F</i> = 0.02; <i>p</i> = 0.8831		
8 th Grade (1 st Year)	KIPP students	9	48.889	13.280	9	46.778	16.445
	Control students	9	49.000	14.027	9	46.000	15.141
8 th Grade (2 nd Year)	KIPP students	3	48.333	15.503	3	51.667	15.503
	Control students	3	48.000	14.526	3	50.667	14.503
8 th Grade (3 rd Year)	KIPP students	3	50.667	26.633	3	54.333	21.362
	Control students	3	51.333	25.403	3	54.667	25.007
8 th Grade (4 th Year)	KIPP students	16	50.563	17.317	16	47.688	13.016
	Control students	16	50.000	17.359	16	46.625	12.727
8 th Grade (Combined)	KIPP students	31	49.871	16.159	31	48.452	14.454
	Control students	31	49.645	16.155	31	47.613	14.319
	Effect Size	<i>d</i> = 0.01			<i>d</i> = 0.06		
	One-way ANOVA	<i>F</i> = 0.00; <i>p</i> = 0.9563			<i>F</i> = 0.05; <i>p</i> = 0.8193		

Table 6. Post-implementation Means, Standard Deviations, and Effect Sizes

Grade and Cohort	Group	Reading/LA				Math			
		<i>N</i>	<i>M</i>	<i>SD</i>	Adj. <i>M</i>	<i>N</i>	<i>M</i>	<i>SD</i>	Adj. <i>M</i>
5 th Grade (1 st Year)	KIPP students	42	33.024	7.754	32.768	42	43.310	6.983	42.975
	Control students	42	34.024	10.084	34.279	42	38.190	10.308	38.525
	Effect Size	$d = -0.11$				$d = 0.59$			
	Adj. Effect Size	$d = -0.28$				$d = 0.74^{**}$			
6 th Grade (1 st Year)	KIPP students	19	35.526	10.715	33.559	19	37.211	10.059	35.423
	Control students	19	33.211	11.033	31.318	19	36.526	10.205	34.817
	Effect Size	$d = 0.22$				$d = 0.07$			
	Adj. Effect Size	$d = 0.34$				$d = 0.10$			
6 th Grade (2 nd Year)	KIPP students	28	35.143	9.344	36.363	28	38.036	9.686	39.176
	Control students	28	31.250	9.571	32.649	28	35.036	9.426	36.268
	Effect Size	$d = 0.42$				$d = 0.32$			
	Adj. Effect Size	$d = 0.57$				$d = 0.46$			
6 th Grade (Combined)	KIPP students	47	35.298	9.809	34.961	47	37.702	9.738	37.300
	Control students	47	32.043	10.117	31.984	47	35.638	9.667	35.543
	Effect Size	$d = 0.33$				$d = 0.22$			
	Adj. Effect Size	$d = 0.44$				$d = 0.27$			
7 th Grade (1 st Year)	KIPP students	12	32.333	10.966	32.174	12	34.167	8.430	34.835
	Control students	12	30.417	8.118	30.264	12	36.417	9.672	36.869
	Effect Size	$d = 0.21$				$d = -0.26$			
	Adj. Effect Size	$d = 0.29$				$d = -0.29$			
7 th Grade (2 nd Year)	KIPP students	7	31.714	9.394	33.083	7	33.714	7.204	34.031
	Control students	7	29.857	7.946	31.095	7	30.714	10.766	31.016
	Effect Size	$d = 0.23$				$d = 0.35$			
	Adj. Effect Size	$d = 0.31$				$d = 0.45$			
7 th Grade (3 rd Year)	KIPP students	26	31.885	9.378	31.692	26	35.038	9.652	34.817
	Control students	26	34.000	7.689	33.635	26	34.808	10.756	34.346
	Effect Size	$d = -0.25$				$d = 0.02$			
	Adj. Effect Size	$d = -0.29$				$d = 0.07$			
7 th Grade (Combined)	KIPP students	45	31.978	9.598	32.316	45	34.600	8.835	34.561
	Control students	45	32.400	7.892	31.665	45	34.600	10.404	34.077
	Effect Size	$d = -0.05$				$d = 0.00$			
	Adj. Effect Size	$d = 0.08$				$d = 0.06$			

Table 6, continued.

Grade and Cohort	Group	Reading/LA				Math			
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>Adj. M</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Adj. M</i>
8 th Grade (1 st Year)	KIPP students	9	35.444	9.015	35.944	9	34.889	11.439	35.466
	Control students	9	35.778	10.402	36.547	9	33.556	11.620	34.476
	Effect Size	<i>d</i> = - 0.04				<i>d</i> = 0.12			
	Adj. Effect Size	<i>d</i> = - 0.1				<i>d</i> = 0.13			
8 th Grade (2 nd Year)	KIPP students	3	38.667	6.658	37.462	3	43.333	4.041	41.745
	Control students	3	39.667	10.970	38.838	3	38.333	10.693	37.197
	Effect Size	<i>d</i> = - 0.13				<i>d</i> = 0.76			
	Adj. Effect Size	<i>d</i> = - 0.25				<i>d</i> = 0.70			
8 th Grade (3 rd Year)	KIPP students	3	38.000	1.732	35.705	3	44.000	5.292	41.175
	Control students	3	34.000	15.395	31.545	3	38.333	15.308	35.346
	Effect Size	<i>d</i> = 0.45				<i>d</i> = 0.61			
	Adj. Effect Size	<i>d</i> = 0.75				<i>d</i> = 0.89			
8 th Grade (4 th Year)	KIPP students	16	35.313	8.154	35.386	16	38.313	9.884	38.449
	Control students	16	31.500	8.524	31.985	16	33.500	9.899	34.122
	Effect Size	<i>d</i> = 0.47				<i>d</i> = 0.50			
	Adj. Effect Size	<i>d</i> = 0.52				<i>d</i> = 0.56			
8 th Grade (Combined)	KIPP students	31	35.935	7.716	36.124	31	38.355	9.796	39.209
	Control students	31	33.774	9.821	34.729	31	34.452	10.582	35.285
	Effect Size	<i>d</i> = 0.25				<i>d</i> = 0.39			
	Adj. Effect Size	<i>d</i> = 0.16				<i>d</i> = 0.39			

* $p < .05$; ** $p < .01$

As Table 7 illustrates, correlations between baseline and post-implementation achievement scores for each grade by subject area and cohort were moderate to strong, ranging from $r = 0.527$ to 0.752 in Math and 0.466 to 0.876 in Reading/LA.

Table 7. Correlations between Pre- and Post-implementation Achievement Scores

Grade and Cohort	Subject	Baseline	Post-implementation	r	p
5 th Grade (1 st Year)	Reading/LA	2004-05 TCAP # Correct	2005-06 TCAP # Correct	0.790	<.0001
	Math	2004-05 TCAP # Correct	2005-06 TCAP # Correct	0.678	<.0001
6 th Grade (1 st Year)	Reading/LA	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.756	<.0001
	Math	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.752	<.0001
6 th Grade (2 nd Year)	Reading/LA	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.685	<.0001
	Math	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.689	<.0001
6 th Grade (Combined)	Reading/LA	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.712	<.0001
	Math	2003-04 TCAP # Correct	2005-06 TCAP # Correct	0.710	<.0001
7 th Grade (1 st Year)	Reading/LA	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.466	0.0217
	Math	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.711	<.0001
7 th Grade (2 nd Year)	Reading/LA	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.876	<.0001
	Math	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.703	0.005
7 th Grade (3 rd Year)	Reading/LA	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.546	<.0001
	Math	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.667	<.0001
7 th Grade (Combined)	Reading/LA	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.576	<.0001
	Math	2002-03 TCAP # Correct	2005-06 TCAP # Correct	0.675	<.0001
8 th Grade (1 st Year)	Reading/LA	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.489	0.0396
	Math	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.682	0.0018
8 th Grade (2 nd Year)	Reading/LA	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.708	0.1156
	Math	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.636	0.1749
8 th Grade (3 rd Year)	Reading/LA	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.715	0.11
	Math	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.527	0.2829
8 th Grade (4 th Year)	Reading/LA	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.603	0.0003
	Math	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.693	<.0001
8 th Grade (Combined)	Reading/LA	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.566	<.0001
	Math	2001-02 TCAP # Correct	2005-06 TCAP # Correct	0.668	<.0001

Post-implementation Outcomes for 5th Grade

MANCOVA indicated a significant effect for KIPP membership (Wilks' Lambda = 0.818, $F(2,79) = 8.77$, $p = 0.0004$). Follow up univariate ANCOVAs indicated a significant effect for KIPP membership in Math ($F(1,80) = 11.22$, $p = 0.0012$), but no statistically significant between-group difference in Reading/LA ($F(1,80) = 1.59$, $p = 0.2110$). The large adjusted effect size in math ($d = 0.74$) illustrates the magnitude of the advantage for KIPP, while the adjusted effect size in Reading/LA ($d = -0.28$) shows that control students tended to outscore KIPP students.

As displayed in Table 9, based on three performance levels (Below, Proficient and Advanced), student proficiency in Math was significantly different between KIPP and control students ($\chi^2 (2) = 9.4095$, $p = 0.0091$), while student achievement in Reading/LA was not ($\chi^2 (2) = 0.6770$, $p = 0.7129$). To examine which pair or pairs of groups had significantly different proportions in Math, standardized Pearson residuals (R) were computed for each pair (cell), and compared to the cut-off 2. If a standardized residual for a cell is greater than 2 (in absolute value), we can conclude that it is a major contributor to the significant chi-square value. As shown in Table 8, the standardized residuals for cell 11 ($R = -2.763$) and cell 21 ($R = 2.763$) were greater than 2.00 (in absolute value), and thus indicated that there were fewer KIPP and more control students than expected who were Below Proficient level. Therefore, the proportions of students at the Below Proficient level were significantly different between KIPP and control students. Using two performance levels (Below and Proficient/Advanced), the chi-square test also indicated a significant difference in Math ($\chi^2 (1) = 7.6364$, $p = 0.0057$) and no difference in Reading/LA ($\chi^2 (1) = 0.0000$, $p = 1.0000$). While 100% of KIPP students were Proficient or

Advanced in Math, only 83.3% of control students met proficiency. Chi-square test results by cohort are shown in Table 9.

Table 8. Standardized Residuals for Follow-up Test for 5th Grade Reading/Language Arts

	Below	Proficient	Advanced
KIPP	-2.763	-0.436	1.992
Control	2.763	0.436	-1.992

Post-implementation Outcomes for 6th Grade

MANCOVA indicated no significant effect for KIPP membership (Wilks' Lambda = 0.949, $F(2,87) = 2.34$, $p = 0.1023$), cohort (Wilks' Lambda = 0.956, $F(2,87) = 1.99$, $p = 0.1430$), or the interaction between KIPP membership and cohort (Wilks' Lambda = 0.992, $F(2,87) = 0.37$, $p = 0.6906$). Although not significant, the adjusted effect sizes in Reading/LA were consistently large across both cohorts (Cohort 1: $d = 0.34$; Cohort 2: $d = 0.57$) and the combined cohorts ($d = 0.44$), and in favor of KIPP students. In Math, the effect size for Cohort 2 ($d = 0.46$) was also very large and in favor of KIPP students.

Based on three performance levels (Below, Proficient and Advanced), no significant differences in achievement were found (see Table 9) in either subject area between KIPP and control students (Reading/LA: $\chi^2(2) = 1.3470$, $p = 0.5099$, and Math: $\chi^2(2) = 0.9761$, $p = 0.6138$). With two performance levels (Below and Proficient/Advanced), no significant differences were found in either Reading/LA or Math ($\chi^2(1) = 1.0162$, $p = 0.3134$, and $\chi^2(1) = 0.9266$, $p = 0.3357$, respectively). KIPP students had a higher percentage scoring Proficient and Advanced compared to controls in both Reading/LA (82.98% vs. 74.47% respectively) and Math (91.49% vs. 85.11% respectively).

Post-implementation Outcomes for 7th Grade

MANCOVA demonstrated no significant effect for KIPP membership (Wilks' Lambda = 0.998, $F(4,62) = 0.08$, $p = 0.9248$), cohort (Wilks' Lambda = 0.953, $F(4,62) = 0.99$, $p = 0.4168$), or the interaction of KIPP membership and cohort (Wilks' Lambda = 0.951, $F(4,62) = 1.03$, $p = 0.3942$).

Based on three performance levels, no significant differences were found (see Table 9) in student achievement in either Reading/LA or Math between KIPP and control students ($\chi^2(2) = 0.8352$, $p = 0.6586$, and $\chi^2(2) = 0.0747$, $p = 0.9633$ respectively). Using two performance levels, the chi-square test also indicated no significant differences in Reading/LA or Math ($\chi^2(1) = 0.8092$, $p = 0.3684$, and $\chi^2(1) = 0.0000$, $p = 1.0000$, respectively). While controls had a higher percentage than KIPP students scoring Proficient and Advanced in Reading/LA (88.89% vs. 82.22% respectively), both groups were equal in the percentage scoring Proficient and Advanced in Math (88.89%).

Post-implementation Outcomes for 8th Grade

MANCOVA indicated no significant effect for KIPP membership (Wilks' Lambda = 0.955, $F(2,51) = 1.2$, $p = 0.3103$), cohort (Wilks' Lambda = 0.875, $F(2,51) = 1.17$, $p = 0.3259$), or the interaction of KIPP membership and cohort (Wilks' Lambda = 0.962, $F(2,51) = 0.33$, $p = 0.9182$). The adjusted effect sizes, while not significant, were consistently large in Math and favored KIPP students. Adjusted effect sizes for Cohort 3 ($d = 0.75$) and Cohort 4 ($d = 0.52$) in Reading/LA were also large and in favor of KIPP students.

Based on three performance levels, chi square (see Table 9) showed no significant differences in student achievement in either subject area (Reading/LA: $\chi^2(2) = 5.1711$, $p = 0.0754$, and Math: $\chi^2(2) = 3.1429$, $p = 0.2077$). Using two performance levels, chi-square

showed a significant difference in Reading/LA that favored KIPP students ($\chi^2 (1) = 5.1667$, $p = 0.0230$), while no significant difference was found in Math ($\chi^2 (1) = 2.2963$, $p = 0.1297$). Over 95% of KIPP students scored Proficient or Advanced in Reading/LA, compared to slightly more than 77% of control students.

Table 9. Chi-square Test Statistics

Grade and Cohort	Subject Area	Group	Below Proficient		Proficient		Advanced		Chi Square (3 levels)	Chi Square (2 levels)
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		
5 th Grade (1 st year)	Reading/LA	KIPP	3	7.14	32	76.19	7	16.67	0.6770	0.0000
		Control	3	7.14	29	69.05	10	23.81		
	Math	KIPP	0	0.00	20	47.62	22	52.38	9.4095**	7.6364**
		Control	7	16.67	22	52.38	13	30.95		
6 th Grade (1 st year)	Reading/LA	KIPP	4	21.05	10	52.63	5	26.32	0.2222	0.1456
		Control	5	26.32	10	52.63	4	21.05		
	Math	KIPP	2	10.53	12	63.15	5	26.32	0.0000	0.0000
		Control	2	10.53	12	63.15	5	26.32		
6 th Grade (2 nd year)	Reading/LA	KIPP	4	14.29	19	67.86	5	17.86	1.3452	1.0182
		Control	7	25.00	18	64.29	3	10.71		
	Math	KIPP	2	7.14	18	64.29	8	28.57	1.5584	1.4694
		Control	5	17.86	15	53.57	8	28.57		
6 th Grade (Combined)	Reading/LA	KIPP	8	17.02	29	61.70	10	21.28	1.3470	1.0162
		Control	12	25.53	28	59.58	7	14.89		
	Math	KIPP	4	8.51	30	63.83	13	27.66	0.9761	0.9266
		Control	7	14.89	27	57.45	13	27.66		

Table 9, continued.

Grade and Cohort	Subject Area	Group	Below Proficient		Proficient		Advanced		Chi Square (3 levels)	Chi Square (2 levels)
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		
7 th Grade (1 st year)	Reading/LA	KIPP	2	16.67	7	58.33	3	25.00	1.2500	0.0000
		Control	2	16.67	9	75.00	1	8.33		
	Math	KIPP	1	8.33	9	75.00	2	16.67	1.0526	1.0435
		Control	0	0.00	10	83.33	2	16.67		
7 th Grade (2 nd year)	Reading/LA	KIPP	1	14.29	5	71.42	1	14.29	0.0000	0.0000
		Control	1	14.29	5	71.42	1	14.29		
	Math	KIPP	0	0.00	6	85.71	1	14.29	2.0000	1.0769
		Control	1	14.29	6	85.71	0	0.00		
7 th Grade (3 rd year)	Reading/LA	KIPP	5	19.23	18	69.23	3	11.54	1.8127	1.4857
		Control	2	7.69	19	73.08	5	19.23		
	Math	KIPP	4	15.38	16	61.54	6	23.08	0.0000	0.0000
		Control	4	15.38	16	61.54	6	23.08		
7 th Grade (Combined)	Reading/LA	KIPP	8	17.78	30	66.66	7	15.56	0.8352	0.8092
		Control	5	11.11	33	73.33	7	15.56		
	Math	KIPP	5	11.11	31	68.89	9	20.00	0.0747	0.0000
		Control	5	11.11	32	71.11	8	17.78		
8 th Grade (1 st year)	Reading/LA	KIPP	1	11.11	5	55.56	3	33.33	0.9762	0.4000
		Control	2	22.22	3	33.33	4	44.44		
	Math	KIPP	1	11.11	6	66.67	2	22.22	0.6667	0.4000
		Control	2	22.22	6	66.67	1	11.11		
8 th Grade (2 nd year)	Reading/LA	KIPP	0	0.00	2	66.67	1	33.33	0.0000	na
		Control	0	0.00	2	66.67	1	33.33		
	Math	KIPP	0	0.00	2	66.67	1	33.33	0.0000	na
		Control	0	0.00	2	66.67	1	33.33		
8 th Grade (3 rd year)	Reading/LA	KIPP	0	0.00	3	100.00	0	0.00	3.0000	1.2000
		Control	1	33.33	1	33.33	1	33.33		
	Math	KIPP	0	0.00	1	33.33	2	66.67	1.3333	1.2000
		Control	1	33.33	1	33.33	1	33.33		
8 th Grade (4 th year)	Reading/LA	KIPP	0	0.00	11	68.75	5	31.25	6.6667*	4.5714*
		Control	4	25.00	11	68.75	1	6.25		
	Math	KIPP	1	6.25	11	68.75	4	25.00	1.6667	1.1429
		Control	3	18.75	11	68.75	2	12.50		
8 th Grade (Combined)	Reading/LA	KIPP	1	3.23	21	67.74	9	29.03	5.1711	5.1667*
		Control	7	22.58	17	54.84	7	22.58		
	Math	KIPP	2	6.45	20	64.52	9	29.03	3.1429	2.2963
		Control	6	19.35	20	64.52	5	16.13		

Note. Chi-square test may not be a valid test for cases that one or more cells have expected counts less than 5.

na Chi-square statistics could not be computed

* $p < .05$; ** $p < .01$

Discussion

A rigorous quasi-experimental research design was employed, in which each KIPP student was individually matched to a highly similar counterpart who attended one of several geographically proximate and demographically similar neighborhood schools. As shown by the pretest analyses, the various fifth- to eighth-grade cohorts of KIPP and control students scored almost identically on the TCAP/AT in the year prior to the former group's enrollment in KIPP. Thus, any differences between groups on the posttest measures would be suggestive of program (KIPP vs. control school) impacts.

Overall, the achievement analyses revealed fairly positive outcomes for KIPP students relative to matched control students. It is important to note that the positive effects were statistically significant (considered "reliable") only for (a) fifth-grade Math in both the percentage of questions answered correctly and in the proportion of students scoring Proficient or Advanced and (b) 8th graders scoring Proficient or Advanced in Reading/LA (in the supplementary analysis). However, out of the 8 overall grade-level analyses comparing KIPP and control students in R/LA and Math (4 grades x 2 subjects), 7 (88%) showed positive effects favoring KIPP. Of note, the median grade-level effect sizes were 0.22 in R/LA and .33 in Math, reflecting educationally moderate to strong advantages (about one-fourth and one-third standard deviation, respectively). Due to the small sample sizes, the power of detecting significant differences was low in some comparisons.

As previously indicated, 2005-06 was a challenging year at KIPP due to leadership discontinuity. The original principal, who was extremely well liked by teachers and parents, unexpectedly retired due to health reasons. He was replaced by interim principals who served until January 2005, when a new principal was hired. This principal had an immediate positive

impact, but then left the school in February 2006. To the extent that strong leadership is critical to school effectiveness, as the educational reform literature clearly suggests (see Datnow, Hubbard, & Mehan, 2002; Desimone, 2002, Rowan et al., 2004), these events most likely disrupted and weakened the implementation of school programs, thereby adversely affecting curriculum, instruction, and student achievement. School climate and teacher attitudes, while still mostly positive, declined relative to prior years. Student behavior was explicitly identified as a very serious problem that had magnified to the point of hindering the quality of instruction. Encouragingly, a “veteran KIPP” principal was hired for the 2006-07 academic year. Teachers are again enthusiastic about the stable leadership. It will be important and revealing to investigate whether, in Year (2006-07), KIPP can demonstrate the potential to raise student achievement evidenced most strongly in Years 1 and 2.

References

- Berends, M., Kirby, S.N., Naftel, S., & McKelvey, C. (2002). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: RAND.
- Bodilly, S. J. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: Rand.
- Bodilly, S. J., & Berends, M. (1999). Necessary district support for comprehensive school reform. In G. Orfield & E. H. DeBray (Eds.), *Hard work for good schools: Facts, not fads, in Title I reform* (pp. 111-119). Cambridge, MA: Harvard University, Civil Rights Project.
- Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Butler, E. D., & Alberg, M. J. (1989). *School Climate Inventory (SCI)*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- CTB/MacMillan/McGraw-Hill (1997). *TerraNova Technical Bulletin I*. Monterey, CA: CTB/McGraw-Hill.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. New York: Routledge-Falmer.
- Edmonds, R. R. (1979, October). Effective schools for the urban poor. *Educational Leadership*, 37, 15-18, 20-24.
- Edmonds, R. R. (1982, December). Programs of school improvement: An overview. *Educational Leadership*, 40, 4-11.
- Frechtling, J., & Sharp, L. (1997). *User-friendly handbook for mixed methods evaluations*. Washington, DC: National Science Foundation.

- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Knowledge is Power Program (KIPP) (n.d.). *About KIPP*. Retrieved November 20, 2005, from <http://www.kipp.org/>.
- Larson, R., Moneta, G., Richards, M., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, 73, 1151-1165.
- Lewis, E. M., Ross, S. M., & Alberg, M. (July, 1999). *School Observation Measure: Reliability analysis*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Miles, M. B., & Huberman, A. M. (Eds.). (1994). *An expanded sourcebook: Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Nunnery, J. A., Ross, S. M., & Sterbinsky, A. (2003). *Constructs underlying teacher perceptions of change in schools implementing comprehensive school reform models*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy. Paper presented at the Annual Meeting of the American Educational Research Association in April, 2003.
- Onwuegbuzie, A., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.). *Handbook of mixed methods in social and behavioral research*. (pp. 351-383). Thousand Oaks, Sage.
- Roeser, R., & Eccles, J. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence*, 8, 123-158.

- Ross, S.M. (December 2001). *Creating Critical Mass for Restructuring: What we can learn from Memphis*. Policy Brief, AEL Organization, Charleston, WV.
- Ross, S. M. (2003). *Effective Schools Correlates as Indicators of Educational Improvement: An Examination of Three Urban Reform Initiatives*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Ross, S. M., & Alberg, M. (1999). *Comprehensive School Reform Teacher Questionnaire*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- Ross, S. M., Henry, D., Phillipsen, L., Evans, K., Smith, L., & Buggey, T. (1997). Matching restructuring programs to schools: Selection, negotiation, and preparation. *School Effectiveness and School Improvement* 8 (1) 45-71
- Ross, S. M., & McDonald, A. J. (2003). *KIPP Parent Questionnaire*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- Ross, S. M., McDonald, A. J., & Alberg, M. (2002). *Program Benchmarking for Raising Student Achievement*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- Ross, S. M., McDonald, A. J., & McSparrin-Gallagher, B. (2004). *Year 1 Evaluation of the KIPP DIAMOND Academy: Analysis of TCAP Scores for Matched Program-Control Group Students*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Ross, S. M., Smith, L. J., & Alberg M. J. (1998). *School Observation Measure*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Ross, S.M., Smith, L.J., Alberg, M., & Lowther, D.L. (2004). "Using Classroom Observation as a Research and Formative Evaluation Tool in Educational Reform: The School

- Observation Measure” in H. Waxman (Ed.). *Observational Research in U.S. Classrooms: New Approaches for Understanding Cultural and Linguistic Diversity* (pp. 144-173). Cambridge, MA: Cambridge University Press.
- Simmons, R., & Blyth, D. (1987). *Moving into adolescence*. New York: Aldine de Gruyter.
- Smith, L., Maxwell, S., Lowther, D., Hacker, D., Bol, L., & Nunnery, J. (1997). Activities in schools and programs experiencing the most, and least, early implementation success. *School Effectiveness and School Improvement, 8*, 125-150.
- Snipes, J. C., & Casserly, M. D. (2004). Urban school systems and education reform: Key lessons from a case study of large urban school systems. *Journal of Education for Students Placed at Risk, 9*(2), 127-142.
- Sterbinsky, A., & Ross, S. M. (2003). *School Observation Measure Reliability Study*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- U.S. Department of Education. (1999). *Guidance on the comprehensive school reform program*. Washington, DC: Author.
- Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher, 32*(7), 22-30.
- Way, N., & Robinson, M. (2003). A longitudinal study of the effects of family, friends, and school experiences on the psychological adjustment of ethnic minority, low-SES adolescents. *Journal of Adolescent Research, 18*, 324-346.

Appendix

Parent Questionnaire Data Summary for Years 1 - 3

Year 1 (2002-2003) N = 29

Year 2 (2003-2004) N = 54

Year 3 (2004-2005) N = 16

Parent Questionnaire Items	Percent Strongly Agree and Agree			Percent Neutral			Percent Disagree and Strongly Disagree		
	1	2	3	1	2	3	1	2	3
1. It was hard for me to adjust to the differences between the Academy and a traditional school.	34.5	33.3	12.5	10.3	14.8	37.5	55.1	51.9	50.1
2. It was hard for my child to adjust to the differences between the Academy and a traditional school.	24.1	44.5	25.1	13.8	14.8	37.5	62.1	40.7	37.5
3. I feel welcome whenever I go to the Academy.	93.1	92.6	100.0	0.0	7.4	0.0	6.8	0.0	0.0
4. The Academy expects students to perform at a high level.	96.5	88.9	93.8	3.4	9.3	6.3	0.0	1.9	0.0
5. My child feels safe at the Academy.	96.6	90.8	87.6	0.0	7.4	12.5	3.4	1.9	0.0
6. I am satisfied with what my child is learning at the Academy.	96.6	87.0	87.5	3.4	3.7	6.3	0.0	5.6	6.3
7. I give my opinions to the Academy to help them plan for the future.	65.5	50.0	50.1	24.1	22.2	31.3	10.3	24.1	18.8
8. I volunteer at the Academy.	55.2	35.2	68.8	27.6	27.8	18.8	17.2	31.5	12.5
9. I participate in parent education and training activities sponsored by the Academy.	65.5	44.5	56.3	20.7	37.0	31.3	13.8	18.5	12.5
10. I participate in special community events sponsored by the Academy.	75.8	57.4	62.6	10.3	25.9	31.3	13.8	16.7	6.3
11. My child's teacher(s) shows me how to help my child learn.	82.7	79.6	87.6	10.3	11.1	0.0	6.8	9.3	12.5
12. My child's teacher(s) always tells me how my child is doing in school.	93.1	85.2	93.8	3.4	7.4	0.0	0.0	7.4	6.3