



**Year 1 Evaluation of the KIPP DIAMOND Academy:
Analysis of TCAP Scores for Matched Program-
Control Group Students**

Steven M. Ross and Aaron McDonald
Center for Research in Educational Policy
The University of Memphis

Brenda McSparrin Gallagher
Office of Research and Evaluation
Memphis City Schools

May 2004

Abstract

In the present study, 49 fifth-graders enrolled the KIPP D.I.A.M.O.N.D. Academy in the 2002-03 school year were individually matched to control student from five feeder schools on the basis of ethnicity, free-reduced lunch status, and fourth-grade achievement on the Reading and Mathematics subtests of the Tennessee Comprehensive Assessment Program: Achievement Test (TCAP:AT). Although the KIPP and control students had virtually identical means on all fourth-grade tests, the KIPP students outperformed the control students on five out of the six fifth-grade tests. The one exception was the Writing assessment, on which KIPP and control group means were virtually identical. In four of the TCAP:AT tests (NRT-Reading, NRT-Math, CRT-Reading/Language Arts, CRT-Math) the comparisons were statistically significant with effect sizes ranging from +0.31 to +0.63. Across all 6 tests, the median adjusted *ES* was +0.31, indicating a moderately strong effect. These positive outcomes compare favorably with those from the most successful whole-school reform models (Borman, Hewes, Overman, & Brown, 2002). Findings are discussed with regard to the KIPP components and conditions likely to have increased academic focus and learning in the first implementation year.

**Year 1 Evaluation of the KIPP D.I.A.M.O.N.D. Academy:
Analysis of Scores on the Tennessee Comprehensive Assessment
Program for Matched Program-Control Group Students**

The present study extends prior evaluation research conducted on the Year 1 evaluation of the KIPP: D.I.A.M.O.N.D. Academy (Alberg, 2003; Ross & Calaway, 2003; Sterbinsky & Ross, 2003). In brief, KIPP:DA is a publicly funded “education choice” school in the Memphis City Schools (MCS) system. Founded in the summer of 2002 and located adjacent to Cypress Middle School, it is governed by the Memphis Board of Education and staffed by MCS employees. However, unlike typical neighborhood schools but similar to many magnet schools throughout the country, KIPP:DA must be chosen by students’ families who, in turn, must agree to abide by school’s expectations for attendance, homework completion, and parent (or caretaker) involvement. KIPP, an acronym for Knowledge is Power Program, is described as an academically rigorous college preparatory program designed to promote high levels of academic achievement and positive student leadership. In Memphis, “Desire, Discipline, and Dedication” are listed as components of the school’s culture, with D.I.A.M.O.N.D. standing for “Daring Individual Achievers Making Outstanding New Dreams.”

The school is in session 7:30 a.m. to 5:00 p.m. during the week, four hours on Saturday, and a month during the summer. Teachers are provided with cellular phones and are available to students and their families outside normal school hours for assistance with homework or in case of emergency. It is important to note that there is no intellectual or documented achievement requirement for admission to KIPP:DA.

However, all students and their parents must sign commitment forms indicating their agreement with the educational mission of the school and their willingness to support the school's rigorous requirements for academic engagement and exemplary conduct.

KIPP: DA teachers receive higher salaries than their peers in MCS because of greater than usual expectations regarding the extended school hours. In addition to the principal, the first-year staff included three full-time teachers, one full-time special education teacher, one part-time speech pathologist, and front office staff. Of the 55 students comprising the three fifth-grade classes, all were (100%) African-American, 60% were female, and 92% were eligible for free or reduced-price lunch.

Prior Research Studies

During the 2002-03 school year, the Center for Research in Educational Policy (CREP) at The University of Memphis performed a "formative evaluation" study of the climate, teaching methods, implementation, and key stakeholder perceptions (i.e., teacher, student, parent, and principal) in the first operational year of the school (Alberg, 2003). Results were generally quite positive, indicating school climate means and teacher reactions above national norms, and substantive progress in implementing programs in curriculum, instruction, and organization. Students and parents also expressed high levels of satisfaction with the school. Teaching methods, however, were predominantly traditional and teacher-centered, raising concerns about maintaining student interest and addressing diverse learning styles throughout the long school day.

In a second study, CREP and the Office of Research and Evaluation (ORE) at Memphis City Schools collaborated to examine the first-year results on the Tennessee Comprehensive Assessment Program (TCAP) writing assessment for KIPP:DA (Ross &

Calaway, 2003). Descriptive findings from longitudinal data showed a gain of +0.28 rubric points for KIPP students, whereas students from a matched control school gained only +0.05 rubric points. Cross-sectional samples revealed a KIPP gain of +0.35 points and a control school decline of -0.18 points. Inferential statistical analysis, however, did not show these patterns to be statistically significant.

In a third study, TCAP: Achievement Test (TCAP:AT) results for KIPP were compared to means for MCS overall and for demographically-comparable neighborhood schools. (Sterbinsky & Ross, 2003). Results indicated that KIPP students actually began the year at an academic deficit compared to the average achievement for Memphis City Schools students in the same grade. However, at the end of the year, KIPP students scored higher than the average MCS 5th grade student in three of the five TCAP:AT subject areas (Social Studies, Math, and Language Arts). In one of the remaining areas (Science), they had reduced the achievement gap, and in the other (Reading) they maintained their relative position. Similarly, after beginning at an academic deficit relative to neighborhood schools, KIPP students scored higher in four of the five TCAP:AT subject areas. Gain scores for KIPP students were noticeably higher than those for: (a) neighborhood schools in all five subject areas, (b) MCS fifth grade students in four of the five TCAP:AT areas, and (c) U.S. norms in three out of the five TCAP:AT subject areas.

Purpose of the Present Study

Results from the above studies, although suggestive about KIPP's generally positive student outcomes, must be interpreted cautiously due to lack of control over sampling selection or other design limitations. Specifically, even though analyses in the

writing assessment study (Ross & Calaway, 2003) were of student-level scores, only one matched control school was involved and all available student data, regardless of the similarity of KIPP and control students, were used. In the TCAP study (Sterbinsky & Ross, 2003), the number of matched control schools was increased to five, but analyses were descriptive comparisons of school-level data only. That is, student-level scores were not yet available. In the present study, we greatly increased the rigor and precision of analyses by establishing treatment-control matches at the individual student level. The primary research questions for the present study were:

1. How did KIPP and control students compare in their reading/language arts, writing, and mathematics achievement on the 2002-03 state assessment?
2. Were results comparable for the norm-referenced and criterion-referenced portions of the test?
3. Were results comparable for different subjects?

Method

Design

A matched treatment-control group design, using 49 student-level matched-pairs, was employed. The potential control group sampling pool was provided by the five elementary schools that feed into KIPP:DA. All were located in the same geographic area, and were highly comparable to KIPP:DA and to each other in student and school demographics. The specific schools and the number of control students selected from each were: Springdale ($n = 8$), Vollentine ($n = 10$), Klondike ($n = 14$), Hollywood ($n = 12$), and Shannon ($n = 5$).

Initially, we identified as possible control group subjects all fifth-grade students enrolled in these schools ($N = 317$) for whom both 2001-02 and 2002-03 scores in reading and mathematics on the TCAP:AT Norm-Referenced Test (NRT) were available. All KIPP:DA and potential control students were African-Americans. We then individually determined the closest individual match for each KIPP:AT student based on (a) gender, (b) poverty (free-reduced lunch status), (c) 2001-02 NRT-Reading subtest, and (d) 2001-02 NRT-Mathematics subtest score. The latter two achievement scores, reflecting student performances in fourth grade, essentially served as a “pretest” or pre-KIPP:DA implementation measure. For the Reading subtest, all matches were within 3 NCE points; for the Mathematics subtest all were within 3 NCE points except for three matched pairs (which were within 5 NCE points).

Descriptive statistics showing KIPP:DA and control means and standard deviations on the pre-implementation measures are provided on Table 1. As can be seen in the table, on all subtests the group means are nearly identical and the associated effect sizes (ES) are close to zero. The absence of any significant group differences was further verified by conducting an ANOVA on each of the measures. All results were nonsignificant: 2001-02 Language, $F(1,96) = .281, p = .597$; Reading, $F = .011, p = .916$; Math, $F = .007, p = .932$.

Student Achievement Measures

The TCAP is a state-mandated standardized system for assessing student achievement in compliance with state policies and the No Child Left Behind (NCLB) act. In 2001-02, TCAP requirements for grades 3-5 consisted of two components—a

multiple-choice NRT and an open-ended writing assessment. In 2002-03, a CRT was added.

Norm-referenced test. 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). For the present study, analyses were restricted to the first three subtests given their direct relevance to research questions and to present NCLB requirements. Normal Curve Equivalents (NCE) were the standardized scores analyzed in this study.

Criterion-referenced test. The purpose of the CRT portion of TCAP:AT is to measure student performance in Reading/Language Arts and Math according to specific 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. For example, for fifth-grade (the present interest), the scale-score criteria in Reading are 621 and 671 for Proficient and Advanced, respectively; these scale scores correspond to number correct cutoffs of 34 and 54, respectively. In Math, the scale-score criteria are 631 (no. correct = 16) and 679 (no. correct = 40) for Proficient and Advanced, respectively.

Writing assessment. The TCAP includes an open-ended Writing assessment, directed by prompts. Starting in 2002-03, the Writing assessment is administered in fifth grade; in prior years, the administration was in fourth grade. Students are asked to write a response to a narrative prompt, the purpose of which is to recount a personal or

fictional experience or tell a story based on a real or imagined event. The students' writing samples are scored by trained judges on a six-point rubric consisting of the following categories: 6 = Outstanding; 5 = Strong; 4 = Competent; 3 = Limited; 2 = Flawed; 1 = Deficient; 0 = Blank, Insufficient, or Off Topic.

Results

Descriptive statistics for post-implementation measures are presented in Table 2. These results show that KIPP:DA students performed directionally higher than control students on all CRT and NRT subtests. Moderate to strong effect sizes, ranging from +0.24 to +0.41, are indicated. Nearly identical KIPP:DA and control group means, however, were obtained on the Writing assessment. Results of inferential analyses are reported in the sections below.

Intercorrelations computed between the four pre-implementation (4th grade) and six post-implementation (5th grade) measures were all statistically significant and at least close to moderate in magnitude (r range = .33 to .91). For example, 5th-grade NRT-Reading correlated with 5th grade NRT-Language Arts at $r = .78$, with 5th grade NRT-Math at $r = .68$, and with NRT-Writing at .57. The 4th grade scores were relatively strong predictors of 5th grade scores as reflected by r 's = .61 for NRT-Language Arts, .71 for NRT-Reading, and .67 for NRT-Math. The 4th and 5th grade Writing scores were moderately correlated at $r = .44$. Finally, NRT and CRT scores were very strongly correlated, with r 's = .88 for Reading and .91 for Math.

2002-03 NRT Language Arts, Reading, and Mathematics

In this analysis, we compared KIPP:DA and control students on the 2002-03 Language Arts, Reading, and Mathematics subtests of the TCAP:AT-NRT. A

multivariate analysis of covariance (MANCOVA), using the 2001-02 (4th grade) Language Arts, Reading, and Mathematics pretest scores as covariates, was conducted. All three covariates were highly significant in the MANCOVA (all p 's < .02). The multivariate effect of Program, however, did not reach significance, $F(3,91) = 2.52$, $p = .063$, $\eta^2 = .077$.

Given the relatively small sample sizes, the a priori hypothesis projecting KIPP:DA advantages, and the approximation to $\alpha = .05$ in the MANCOVA (see Wainer & Robinson, 2003), we proceeded to conduct univariate tests (ANCOVA) on each of the dependent measures. The univariate results were significant for Reading, $F(1,93) = 5.55$, $p = .021$, $\eta^2 = .056$; and Math, $F = 5.74$, $p = .019$, $\eta^2 = .056$; but were nonsignificant for Language Arts, $F = 2.77$, $p = .099$, $\eta^2 = .029$. The adjusted means and associated effect sizes are summarized in Table 2.

2002-03 CRT Reading and Mathematics

An initial analysis examined the percentages of KIPP:DA and control students who scored at Below Proficient, Proficient, and Advanced levels on the CRT Reading/Language Arts and Math subtests. A summary is provided in Table 3. As can be seen, in both Reading/Language Arts and Mathematics, KIPP:DA students were more likely to be represented in the Proficient and Advanced categories than were control students. For example, on the Reading/Language Arts subtest, 10% of KIPP:DA students as compared to 2% of the control students scored at the Advanced level; in Math, the percentages were 16% vs. 0%, respectively. Two-way chi square (Program X Proficiency Level) analyses were significant for Math, $X^2(2) = 8.62$, $p = .013$, but nonsignificant for Reading/Language Arts, $X^2(2) = 4.17$, $p = .124$.

A multivariate analysis of covariance (MANCOVA), using the 2001-02 (4th grade) NRT Language Arts, Reading, and Mathematics subtest scores as covariates, was performed on the 2002-03 CRT Reading and Mathematics subtests. Both the Reading and Math covariates were significant in the MANCOVA (both p 's < .001); the Language Arts covariate, however, was nonsignificant ($p = .44$). Most importantly, the Program effect was highly significant, $F(2,91) = 5.70$, $p = .005$, $\eta^2 = .111$.

Univariate ANCOVAs conducted on the two CRT subtests were significant for both Reading, $F(1,92) = 4.76$, $p = .032$, $\eta^2 = .049$; and Math, $F = 10.82$, $p = .001$, $\eta^2 = .105$. Note from Table 2 that adjustments in the means due to covariate effects increased the effect size favoring KIPP:DA in Reading from +0.28 to +0.31 and in Math from +0.41 to a fairly strong +0.63.

2002-03 Writing Assessment

An examination of the 4th grade Writing levels indicated that 58% of the control students and 50% of the KIPP:DA students scored at levels of Competent or higher. In fifth-grade (2002-03), these percentages improved to 63% and 70%, respectively. However, despite the directional KIPP:DA advantage in 2002-03, chi-square results failed to indicate a significant ($p = .34$) relationship between programs and performance levels.

An analysis of covariance, using the 2001-02 (4th grade) Writing scores as a covariate, was performed on the 2002-03 Writing assessment. Although the covariate was highly significant ($p < .01$), the Program effect was close to zero, $F(1,82) = .004$, $p = .953$, $\eta^2 = .000$. As shown in Table 2, KIPP:DA and control students had almost identical writing means.

Discussion

In the present study, 49 fifth-graders enrolled in the 2002-03 school year were individually matched to control students from five feeder schools on the basis of ethnicity, free-reduced lunch status, and fourth-grade achievement on the Reading and Mathematics subtests of the TCAP:AT. Although the KIPP:DA and control students had virtually identical means on all fourth-grade tests (see Table 1), the KIPP:DA students outperformed the control students on five out of the six fifth-grade tests (see Table 2). The one exception was Writing, on which KIPP:DA and control group means were virtually identical. In four of the TCAP:AT tests (NRT-Reading, NRT-Math, CRT-Reading/Language Arts, CRT-Math) the comparisons were statistically significant with effect sizes ranging from +0.31 to +0.63. Across all 6 tests, the median adjusted *ES* was +0.31, indicating a moderately strong effect.

By comparison, in a recent meta-analytic study of 29 Comprehensive School Reform (CSR) models, Borman, Hewes, Overman, and Brown (2002) found an overall effect size of from +0.10 to +0.14, with the range for the “most successful” category being +0.17 to +0.21. Only 3 out of the 29 models achieved this high status (Direct Instruction, School Development Program, and Success For All). Thus, the KIPP:DA overall results compare favorably to outcomes associated with the usage of other whole-school reform programs. The results are further impressive given the viewpoint by scholars of school reform (Fullan, 2000; Sizer, 1992; Levin, 1993) that school change takes several years to manifest itself in observable outcomes.

Given all the possible variables in a school that can impact educational outcomes, it is highly difficult, if not impossible, for studies of whole-school reform to

provide conclusive evidence about the effects of specific program components (Datnow, Hubbard, & Mehan, 2002; Ross, 2003). To do so, it would be necessary to manipulate different variations of the particular reform approach, examining what occurs when the component of interest (e.g., extended day) is present and absent. In the case of the present research study, KIPP:DA appeared to inherit several possible advantages that might have increased its chances of demonstrating success in Year 1. One factor was the very small school size (only about 50 students). A second was the single grade level established. A third was greater ability than in other district schools to recruit effective teachers. A fourth was extremely high involvement and support by community and university partners. In future years, many of these initial advantages will diminish as the school incorporates sixth- and seventh-grades, increases enrollment, adds teachers, and becomes less novel as a local reform initiative.

These factors notwithstanding, the first-year implementation of KIPP:DA also had to overcome multiple challenges, including sharing a building with another school, enrolling a greater number than expected of special needs students, and preparing teachers, administrators, students, and parents for novel educational structures and events. Given these challenges and the educationally meaningful program effects demonstrated, it certainly seems probable that certain program elements did work to raise student achievement. Although we cannot be certain about the relative contribution of individual elements, we believe that the most influential ones include the extended school day, high parent involvement, positive school climate, and strong commitment by all participant group to high academic rigor.

As the present research on KIPP:DA extends to the second implementation year, it will be both revealing and important to determine whether the positive results of the first year are replicated. If the above program components remain effectual, a likely outcome would be even stronger achievement advantages for KIPP:DA students over their control counterparts when they complete their sixth-grade year.

References

- Alberg, M. (2003). *KIPP D.I.A.M.O.N.D Academy: The first year*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2002). Comprehensive school reform and student achievement: A meta-analysis. *Baltimore, MD: Review of Educational Research, 73(2)*, 125-230.
- CTB/MacMillan/McGraw-Hill (1997). *Terra Nova Technical Bulletin 1*. Monterey, CA: Macmillan/McGraw-Hill.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. New York: NY: Routledge-Farmer.
- Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change, 1(1)*, 5-28.
- Levin, H. (1993). *Learning from Accelerated Schools*. Unpublished paper. Palo Alto, CA: Stanford University.
- Ross, S. M. (2003). Effective schools correlates as indicators of educational improvement: An examination of three urban reform initiatives. *Journal for Effective Schools, 2(2)*, 67-81.
- Ross, S. M., & Calaway, F. (2003). *First-year results on the TCAP writing assessment for the KIPP D.I.A.M.O.N.D. Academy*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Sizer, T.R. (1992). *Horace's school: Redesigning the American high school*. New York: Houghton, Mifflin.
- Sterbinsky, A., & Ross, S. M. (2003). *KIPP D.I.A.M.O.N.D. Academy: Three comparison studies*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher, 32(7)*, 22-30.

Table 1.

Descriptive statistics for KIPP and Control students on 4th grade (pretest) TCAP measures.

Program	2001-02 NRT-Lang. Arts^a	2001-02 NRT-Reading	2001-02 NRT-Math	2001-02 Writing
KIPP:DA				
<i>M</i>	41.16	40.25	41.65	3.48
<i>(SD)</i>	(18.36)	(17.72)	(16.87)	(.976)
Control				
<i>M</i>	43.16	39.92	41.37	3.51
<i>(SD)</i>	(18.61)	(16.45)	(16.10)	(.757)
<i>Effect Size</i>	-0.11	+0.02	+0.02	-0.03

^aNRT= Norm-Referenced portion of the Tennessee Comprehensive Assessment Program: Achievement Test (TCAP:AT).

Note. Both KIPP:DA and control *n*'s = 49 students.

Table 2.

Descriptive statistics for KIPP and Control students on 5th grade (posttest) TCAP measures.

Posttest-Implementation Measures						
Program	2002-03 NRT-Lang. Arts ^a	2002-03 NRT-Reading	2002-03 NRT-Math	2002-03 CRT-Read/ Language Arts ^b	2002-03 CRT-Math	2002-03 Writing
KIPP:DA						
<i>M</i>	42.80	43.08	42.84	641.92	633.82	3.88
<i>M_{adj}</i>	42.97	43.38*	42.87	642.38*	634.28**	3.89
(<i>SD</i>)	(16.01)	(18.47)	(16.42)	(31.59)	(32.94)	(0.86)
Control						
<i>M</i>	38.96	38.18	37.80	633.00	622.15	3.91
<i>M_{adj}</i>	38.81	37.89*	37.76*	632.53*	616.29**	3.90
(<i>SD</i>)	(16.28)	(15.40)	(11.73)	(30.91)	(22.59)	(0.92)
<i>Effect Size</i>	+0.24	+0.29	+0.35	+0.28	+0.41	-0.03
<i>ES_{adj}</i>	+0.26	+0.31	+0.35	+0.31	+0.63	-0.01

* $p < .05$ for KIPP vs. control means in ANCOVA. ** $p < .01$.

^aNRT = Norm-Referenced Test portion of the Tennessee Comprehensive Assessment Program: Achievement Test (TCAP:AT);

^bCRT = Criterion-Referenced Test portion of the TCAP:AT

Table 3.

The percentages of KIPP and control students scoring at different proficiency levels on the CRT-Reading/Language Arts and CRT-Mathematics in Fifth Grade.

Group and Subject	Proficiency Levels		
	Below Proficient	Proficient	Advanced
Reading/Language Arts			
KIPP:DA	35	55	10
Control	50	48	2
Mathematics*			
KIPP:DA	41	43	16
Control	46	54	0

* $p < .05$ in chi-square test