



Analysis of Year 2 (2003 – 2004)

**Student Achievement Outcomes for the
Memphis KIPP DIAMOND Academy**

Brenda McSparrin Gallagher
Office of Research and Evaluation
Memphis City Schools

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

January, 2005

**Analysis of Year 2 (2003 – 2004) Student Achievement Outcomes
for the Memphis KIPP DIAMOND Academy**

Abstract

The present study examined outcomes on the Tennessee Comprehensive Assessment Program/Achievement Test (TCAP/AT) for the KIPP:DIAMOND Academy (KIPP:DA), which is in its second year of operation. Importantly, as in Year 1, a rigorous quasi-experimental research design was employed, in which each KIPP:DA student was individually matched to a highly similar counterpart who attended one of several geographically proximate neighborhood schools. The 2004 TCAP/AT was taken by three subsamples of KIPP:DA students: (a) fifth graders who had completed their first year at the school, (b) sixth graders who had also completed their first year, and (c) sixth graders who had completed their second year at the school. “Pretest” analyses revealed that KIPP:DA and control students scored almost identically on the TCAP/AT in the year prior to the former group’s enrollment in KIPP:DA.

Comparisons between the three KIPP:DA groups and their matched control counterparts on each of the TCAP/AT subtests (Reading, Language Arts, Mathematics), directionally favored KIPP:DA in 7 out of the 9 analyses (median $ES = +0.14$) unadjusted for pretest score, and in 8 out of 9 analyses (median $ES = +0.16$) adjusted for pretest scores. Inferential analyses showed the KIPP:DA advantages to be statistically significant for only the fifth-grade and sixth-grade longitudinal cohort subsamples. Supplementary descriptive analyses of the criterion-referenced (CRT) portion of TCAP/AT further revealed that, while not statistically significant, KIPP:DA students in all three subsamples were more frequently represented than the matched

control students at the Proficient and Advanced levels, and less frequently at the Below Proficient level, in both Reading/Language Arts and Mathematics. These results are clearly suggestive of positive KIPP:DA effects in Year 2, especially in view of the doubling of school size and special unanticipated challenges faced during the year.

Analysis of Year 2 (2003 – 2004) Student Achievement Outcomes for the Memphis KIPP DIAMOND Academy

The Year 1 (2002-03) evaluation of the Memphis KIPP DIAMOND Academy showed positive implementation and achievement outcomes (Alberg, 2003; Ross, McDonald, & Gallagher, 2004). In that year, only the fifth-grade was established, with an enrollment of approximately 50 students. In Year 2 (2003-04), a sixth-grade was added, thus approximately doubling the enrollment. Similar to the first year, the Year 2 evaluation consisted of two major studies: (a) quantitative analyses of student achievement on the Tennessee Comprehensive Assessment Program: Achievement Test (TCAP/AT) by KIPP students compared to individually-matched control students and (b) a descriptive analysis of program implementation, participant and stakeholder perceptions, and school climate (McDonald et al., in press). The present report is of the student achievement study only.

To provide a comparative framework for interpreting the Year 2 results, a brief review of the first-year findings seems relevant. (The detailed report can be found in Ross et al., 2004). For the Year 1 study, KIPP:DA fifth-grade students ($n = 49$) 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, the KIPP:DA fifth-grade students outperformed the control students on five out of the six tests. The one exception was Writing, on which KIPP:DA and control group means were virtually identical. In four of the TCAP/AT norm-referenced (NRT) and-criterion-referenced (CRT) tests, the

comparisons were statistically significant (NRT-Reading, NRT-Math, CRT-Reading/Language Arts, CRT-Math) with effect sizes (ES) ranging from +0.31 to +0.63. Across all 6 tests, the median adjusted *ES* was +0.31, indicating a moderately strong effect.

The first-year achievement results were thus clearly positive, despite the challenges of establishing a new school in which all KIPP:DA students but relatively few control students were transfers. The descriptive study reinforced and helped to explain these positive outcomes by showing high levels of satisfaction with KIPP:DA by teachers, students, and parents (Alberg, 2003). School climate was extremely positive, with the seven climate dimension means all far exceeding national norms. Teacher methods, while somewhat more teacher-centered (e.g., using direct instruction) than program goals encouraged, revealed generally high academic focus and student engagement.

Evaluation Purpose and Context

In view of the Year 1 outcomes and the changes in KIPP:DA due to its expansion in size and prior year of implementation experience, determining its success in raising student achievement in Year 2 naturally acquires interest and importance. For interpreting the results, an event that appeared to impact the school significantly throughout the year should be noted (McDonald et al., in press). Early in the fall, the original KIPP:DA principal became ill and was on sick leave for many days. By mid-year, he resigned his position and was replaced by an interim principal from the national organization. As described in the implementation study (McDonald et al., in press), the

apparent impact was reducing faculty morale and impeding progress in implementing school programs.

Methodology

Achievement Measures

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 will be administered.

Participants

Participants in the study were both new and continuing KIPP:DA students. The new 2003-04 students were 51 fifth graders and 17 sixth graders. Each student had scores on both the spring, 2003 (pre-implementation or pretest) and spring, 2004 (posttest) TCAP/AT. Continuing students were the 40 sixth graders who were enrolled in KIPP:DA during both the 2002-03 and the 2003-04 school years. Eight students out of the 49 students from the original fifth-grade cohort did not re-enroll at KIPP:DA for the 2003-04 school year. One control student original match did not re-enroll in Memphis City Schools. Each of the continuing students had TCAP/AT scores for spring, 2002 (pretest), spring, 2003 (Year 1 posttest), and spring 2004 (Year 2 posttest). Normal Curve Equivalent (NCE) scores for the Reading, Language Arts, and Mathematics subtests were used in all analyses.

Each KIPP:DA student was matched with a control student based on six criteria: grade level, race, gender, and TCAP/AT pretest scores in Reading, Language Arts, and Mathematics. The possible control group subjects were enrolled in one of the five elementary schools that feed into KIPP:DA (see Table A.1 in the appendix). All five schools were located in the same geographic area and were highly comparable to KIPP:DA and each other in student and school demographics. Fifth graders and sixth graders who were new to KIPP:DA in the 2003-04 school year were matched on the previous year's (spring, 2003) TCAP/AT scores. Possible control students for these students had to have both spring, 2003 and spring, 2004 TCAP/AT scores available. The returning sixth grade KIPP cohort students had been previously matched on spring, 2002 TCAP/AT scores.

Results

Pre-Implementation

Descriptive Results

Descriptive statistics for the pre-implementation TCAP/AT for KIPP:DA and the control students are presented in Table 1. In the year prior to the KIPP:DA subsamples' enrollment in KIPP:DA, all groups scored below the national norm ($M = 50$) on all three subtests (Reading, Language Arts, and Mathematics). With two exceptions, all of the associated effect sizes (ES) comparing KIPP:DA and control students' pretest means were extremely small and close to zero, suggesting a high degree of comparability. The exceptions were weak to moderate advantages for the sixth-grade longitudinal cohort control group in Mathematics ($ES = -0.16$), and the new sixth-grade KIPP:DA students in Reading (+0.13).

Table 1**Descriptive Statistics for KIPP:DA and Control Students on Pretest* NRT TCAP Measures**

Group		Reading	Language Arts	Mathematics
KIPP:DA				
2003-04 New 5th Graders <i>n</i> =51	Mean	39.37	42.69	39.08
	Std. Dev	(15.62)	(15.91)	(13.52)
Control Group <i>n</i> =51	Mean	38.57	42.45	38.29
	Std. Dev	(15.18)	(16.28)	(12.78)
	Effect Size	+0.05	+0.01	+0.06
KIPP:DA				
2003-04 New 6th Grades <i>n</i> =17	Mean	45.53	44.06	40.24
	Std. Dev	(16.07)	(15.73)	(11.01)
Control Group <i>n</i> =17	Mean	43.53	44.65	40.94
	Std. Dev	(15.33)	(15.60)	(9.91)
	Effect Size	+0.13	-0.04	-0.07
KIPP:DA				
2002-03 Cohort 6th Graders <i>n</i> =40	Mean	41.18	41.60	42.63
	Std. Dev	(18.88)	(19.60)	(18.17)
Control Group <i>n</i> =40	Mean	40.70	44.63	42.40
	Std. Dev	(17.48)	(19.30)	(17.12)
	Effect Size	+0.03	-0.16	+0.01

*Pretest is Spring 2003 Scores for the KIPP 2003-04 New 5th and New 6th Graders and Spring 2002 for the 2002-03 KIPP Cohort Group.

Inferential Results

Separate one-way ANOVAs comparing the KIPP:DA and control groups in each subsample yielded nonsignificant results on all three subtests (Reading, Language Arts, and Mathematics). Thus, there is no basis for inferring that either KIPP:DA or control students had an achievement advantage over the other in the year prior to KIPP:DA enrollment.

Post-Implementation

Correlational Results

Correlations between the pre- and post-implementation achievement subtests for the 2003-04 5th grade KIPP:DA and control students combined were all at least moderate in strength and significant at the .01 level: Reading, $r = +.58$, Language Arts, $r = +.63$, and Mathematics, $r = +.56$. Intercorrelations between subject area subtests were all significant and ranged from $+0.49$ to $+0.83$.

Pre- and post-implementation correlations for the 2003-04 new sixth-grade KIPP:DA and control students were all moderate to strong and significant at the .01 level: Reading, $r = +.69$, Language Arts, $r = +.64$, and Mathematics, $r = +.74$. Intercorrelations between subject tests were significant and ranged from $+0.45$ to $+0.86$.

Pre- and post-implementation Year 1 (2002-03) correlations for the 6th grade longitudinal cohort group were all moderate to strong and significant at the .01 level: Reading, $r = +.71$, Language Arts $r = +.65$, and Mathematics, $r = +.67$. Pre- and post-implementation Year 2 (2003-04) correlations were all strong and also significant at the .01 level: Reading, $r = +.70$, Language Arts, $r = +.73$, and Mathematics, $r = +.77$. Intercorrelations between subject area tests ranged from $+0.58$ to $+0.82$. All intercorrelations were significant at the 0.01 level.

Descriptive Results

Descriptive statistics for post-implementation measures are presented in Table 2. All means still fell below the national norm of 50. Inspection of the unadjusted means¹ shows that the 2003-04 KIPP:DA 5th graders directionally surpassed the control students on all three subtests, with *ESs* ranging from $+0.08$ to $+0.35$. For the 2003-04

¹ These are the “obtained” or actual means, unadjusted for differences in pre-implementation scores.

new 6th graders, KIPP:DA students directionally surpassed the control students only in Mathematics ($ES = +0.14$), while showing a moderate deficit in Language Arts ($ES = -0.35$). The 2002-03 KIPP:DA longitudinal cohort group directionally surpassed the control group in Reading ($ES = +0.27$), Language Arts ($ES = +0.31$), and Mathematics ($ES = +0.31$).

Table 2
Descriptive Statistics for KIPP:DA and Control Students on Posttest (Spring, 2004) NRT TCAP Measures

Group		Reading	Language Arts	Mathematics
Year 2 (2003-04):				
KIPP:DA 2003-04 New 5th Grade <i>n</i> =51	Mean	40.57	43.14	37.08
	Adj. Mean	40.50	43.06	37.00
	Std. Dev	(15.42)	(16.78)	(15.93)
Control Group <i>n</i> =51	Mean	39.29	36.92	35.33
	Adj. Mean	39.36	37.00	35.41
	Std. Dev	(16.61)	(17.63)	(15.37)
	Effect Size	+0.08	+0.35	+0.11
	Adj. Effect Size	+0.07	+0.34	+0.10
Year 2(2003-04)				
KIPP:DA 2003-04 New 6th Grade <i>n</i> =17	Mean	39.06	42.71	38.41
	Adj. Mean	39.28	42.86	38.58
	Std. Dev	(17.28)	(13.98)	(20.12)
Control Group <i>n</i> =17	Mean	39.12	46.82	36.12
	Adj. Mean	38.90	46.67	35.95
	Std. Dev	(12.29)	(11.83)	(16.64)
	Effect Size	0.00	-0.35	+0.14
	Adj. Effect Size	+0.03	-0.32	+0.16
Year 1 (2002-03)				
KIPP:DA 2002-03 Cohort 6th Grade <i>n</i> =40	Mean	42.6	42.85	43.65
	Adj. Mean	43.63	43.71	44.39
	Std. Dev	(19.92)	(16.80)	(17.61)
Control Group <i>n</i> =40	Mean	39.68	40.5	38.43
	Adj. Mean	38.65	39.65	37.69
	Std. Dev	(15.65)	(16.39)	(12.59)
	Effect Size	+0.19	+0.14	+0.41
	Adj. Effect Size	+0.32	+0.25	+0.53
Year 2 (Spring 2004)				
KIPP 2002-03 Cohort 6th Graders <i>n</i> =40	Mean	38.88	43.98	39.08
	Adj. Mean	39.71	45.06	40.02
	Std. Dev	(15.50)	(19.40)	(18.28)
Control Group <i>n</i> =40	Mean	34.95	38.53	33.98
	Adj. Mean	34.11	37.45	33.03
	Std. Dev	(14.56)	(17.38)	(16.53)
	Effect Size	+0.27	+0.31	+0.31
	Adj. Effect Size	+0.39	+0.44	+0.42

Inferential Results

NRT Reading, Language Arts, and Mathematics subtests. For the 2003-04 new KIPP:DA fifth- and sixth-grade groups, a multivariate analysis of covariance (MANCOVA) was used to assess program effects. Each KIPP:DA group and its matched control group were separately compared on the 2004 NRT TCAP/AT Reading, Language Arts, and Mathematics subtests. The spring 2003 (pretest) TCAP/AT Language Arts subtest was used as a covariate due to its having the strongest correlation with the 2004 (posttest) scores.

The MANCOVA performed on the fifth-grade KIPP:DA and control group scores yielded a highly significant 2003 Language Arts covariate effect ($p < .001$). The multivariate effect of Program (KIPP:DA vs. control), however, did not reach significance, $F(3,97)=2.27$, $p=.084$. Given the approximation to alpha = .05 in the MANCOVA, we proceeded to conduct univariate tests (ANCOVA) on each of the dependent measures. The univariate results were significant for Language Arts, $F(1,99)=5.25$, $p = .024$; showing an advantage ($ES = +0.33$) for KIPP:DA students ($M_{adj} = 43.06$) over the control group ($M_{adj} = 37.00$) (see Table 2). No differences, however, were indicated for Reading, $F(1,99)=.20$, $p=.653$; or Mathematics, $F(1,99)=.44$, $p=.510$.

For the new sixth-grade KIPP:DA subsample, the 2003 Language Arts covariate was highly significant in the MANCOVA ($p<.001$). The multivariate effect of program, however, did not reach significance, $F(3,29)=1.36$, $p=.273$. It is important, however, to consider that the very small sample size (17 matched pairs) limits the power of the analysis for detecting program effects.

For the longitudinal sixth-grade cohort, a repeated-measures multivariate analysis of covariance (MANCOVA) was used to assess program effects. KIPP:DA and matched control groups were compared on the three TCAP/AT subtests (see Table 2) in 2003 (program Year 1) and 2004 (Year 2). The spring 2002 (pretest) TCAP/AT Language Arts subtest was used as a covariate. The within-subjects factors were subtest (Reading, Language Arts, and Mathematics) and year (2003 and 2004). The between-subjects factor was program (KIPP:DA or control group).

The 2002 Language Arts covariate was highly significant ($p < .001$). Of note, the multivariate between-subjects main effect for program was also significant, $F(3,75)=3.07$, $p=.033$; as was the within-subjects main effect for year, $F(3,75)=4.81$, $p=.004$. The program X year interaction was not significant, $F(3,75)=.51$, $p=.680$, thereby failing to qualify the main effects. That is, a significant interaction would suggest that KIPP:DA vs. control group differences in achievement varied from 2003 to 2004.

Pairwise follow-up comparisons for the year main effect were significant for Reading ($p=.004$) and Mathematics ($p=.001$), but not for Language Arts ($p = .775$). For both of the former subtests, means were higher in 2003 than in 2004. Because these differences reflect combined KIPP:DA and control group scores, and may be due to differential test difficulty across years, they have limited importance for the present study of program impacts.

Means involved in the significant program main effect are presented in Table 3. The univariate follow-up comparisons were significant for all three subtests: Reading, $F(1,77)=6.22$, $p=.015$, Language Arts, $F(1,77)=6.23$, $p=.015$, and Mathematics,

$F(1,77)=7.83$, $p=.006$. KIPP:DA students had higher means than the control group students.

Table 3

Subtest Means for 2003 and 2004 Combined by Program Group on the NRT TCAP/AT

Subtest	KIPP:DA (<i>n</i> = 40)	Control (<i>n</i> = 40)	Difference	Significance
Reading	41.67	36.38	+5.29	0.015
Language Arts	44.38	38.55	+5.83	0.015
Mathematics	42.20	35.36	+6.84	0.006

CRT Reading/Language Arts and Mathematics

A supplementary analysis was performed to examine the percentages of KIPP:DA and control students who scored at Below Proficient, Proficient, and Advanced levels on the 2004 CRT Reading/Language Arts and Mathematics subtests. A summary is provided in Table 4. On both subtests, KIPP:DA students were more likely to be represented in the Proficient and Advanced categories than were control students. In every case the control group had a higher percentage of students scoring in the Below Proficient category. For example, for the sixth-grade longitudinal cohort, 45% and 50% of the control group scored Below Proficient in Reading/Language Arts and Mathematics, respectively, compared to only 30% of the KIPP:DA group in both subjects. Two-way chi square (Program X Proficiency Level) analyses were significant for the KIPP sixth-grade longitudinal cohort and control groups for Mathematics in 2003,

$\chi^2(2)=9.25$, $p=.010$. All other chi square analyses were nonsignificant. Thus, these directional trends, while favorable for KIPP:DA, need to be viewed cautiously.

Table 4

Percent of KIPP and Control Students at Different Proficiency Levels on the Spring, 2004 CRT TCAP

Group	Subject	Below Proficient	Proficient	Advanced	Proficient and Advanced
2003-04 New KIPP:DA 5th Graders $n=51$	2004 Reading/ Language Arts	22	71	8	79
	2004 Mathematics	33	61	6	67
Control Group $n=50$	2004 Reading/ Language Arts	34	60	6	66
	2004 Mathematics	36	54	10	64
2003-04 New KIPP:DA 6th Graders $n=16$	2004 Reading/ Language Arts	25	69	6	75
	2004 Mathematics	25	63	13	76
Control Group $n=17$	2004 Reading/ Language Arts	29	71	0	71
	2004 Mathematics	47	53	0	53
2002-03 KIPP:DA 6th Grade Cohort $n=40$	2004 Reading/ Language Arts	30	63	8	71
	2004 Mathematics	30	53	18	71
	2003 Reading/ Language Arts	33	55	13	68
	2003 Mathematics	40	40	20	60
Control Group $n=39$	2004 Reading/ Language Arts	45	48	8	56
	2004 Mathematics	50	38	13	51
	2003 Reading/ Language Arts	46	51	3	54
	2003 Mathematics	41	59	0	59

Conclusions

Second-year student achievement results for KIPP:DA were positive overall, thereby corroborating and extending the favorable outcomes from Year 1 (Ross et al., 2004). Importantly, as in Year 1, a rigorous quasi-experimental research design was employed, in which each KIPP:DA student was individually matched to a highly similar counterpart who attended one of several geographically proximate neighborhood schools. As shown by the pretest analyses, the KIPP:DA and control students scored almost identically on the TCAP/AT in the year prior to the former group's enrollment in KIPP:DA. Thus, any differences between groups on the posttest measures would certainly appear suggestive of program (KIPP:DA vs. control school) impacts.

The 2004 TCAP/AT was taken by three subsamples of KIPP:DA students: (a) fifth graders who had completed their first year at the school, (b) sixth graders who had also completed their first year, and (c) sixth graders who had completed their second year. Comparisons between these three groups and their matched control counterparts on each of the TCAP/AT subtests, directionally favored KIPP:DA in 7 out of the 9 analyses (median $ES = +0.14$) unadjusted for pretest score, and in 8 out of 9 analyses (median $ES = +0.16$) adjusted for pretest scores. These descriptive results are noteworthy considering that they compare favorably to the overall effect sizes ranging from +0.10 to +0.14 reported by Borman et al. (2003) in a meta-analysis of research on Comprehensive School Reform (CSR) models. The CSR models provided research-based practices and intensive professional development, and, in the vast majority of cases, were supported by federal grants of at least \$50,000 for each of three years.

Inferential statistical analyses, however, showed the KIPP:DA advantages to be significant (reliably greater than chance) for only the fifth-grade and the sixth-grade longitudinal cohort subsamples. Although the new sixth-grade subsample was small and thus subject to sampling error, it seems logical that students transferring to a new school to join an existing cohort of classmates might experience greater adjustment difficulties than would other groups. However, even these new sixth-graders demonstrated higher levels of proficiency on the CRT in both Reading/Language Arts (75% Proficient or Advanced) and Mathematics (76%) than did their control counterparts (71% and 53%, respectively).

As previously indicated, 2003-04 was a challenging year for KIPP:DA stemming from the sudden retirement of its principal and doubling of its size through the expansion to sixth grade. Not surprisingly, school climate and program implementation were weaker than in Year 1 (McDonald et al., in press). Yet, the present fifth-grade and notably, sixth-grade longitudinal cohort analyses essentially replicate the favorable Year 1 results. It will be important and revealing to continue to assess implementation and achievement outcomes as KIPP:DA expands to grade 7 in 2004-05.

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. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- McDonald, A. J, Thompson, S., & Thompson, V. L. (in press). KIPP DIAMOND Academy 2003-2004 Evaluation Report. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- Ross, S. M., McDonald, A. J., & Gallagher, B. M. (2004). *Year 1 Evaluation of the KIPP DIAMOND Academy: Analysis of TCAP Scores for Matched Program-Control Group Students*. Center for Research in Educational Policy, The University of Memphis.

Appendix

Table A-1
Pretest School Attendance for Control Group Matches

KIPP:DA Control Groups	Shannon	Springdale	Vollentine	Klondike	Hollywood
New 5th Grade Control Group Number of Students (percent)	10 (20%)	8 (16%)	13 (25%)	12 (24%)	8 (16%)
New 6th Grade Control Group Number of Students (percent)	3 (18%)	3 (18%)	3 (18%)	5 (29%)	3 (18%)
6th Grade Cohort Control Group Number of Students (percent)	5 (13%)	7 (18%)	8 (20%)	11 (28%)	9 (23%)