Evaluating Success:
KIPP Educational Program Evaluation

Prepared by

New American Schools
Education Performance Network

Technical Report
October 2002

Harold C. Doran
Darrel W. Drury

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Executive Summary

Overview

This report, prepared by the Education Performance Network (EPN) at New American Schools (NAS), presents the findings of an independent evaluation of three new KIPP charter schools. The primary purpose of the analysis was to determine whether or not these schools, in their first year of operation, were able to replicate the academic success of the two original KIPP schools in Houston and New York, as measured by student achievement gains.

The results of this evaluation provide evidence that students' test scores improved at impressive rates after their enrollment in the KIPP schools. Of critical importance, these gains were reflected across demographic subgroups and exceeded those achieved by these same students in the year prior to their enrollment.

The achievement gains in all schools—and across all subtests—were statistically significant, providing evidence of the effectiveness of the KIPP educational program. In addition, students participating in state tests—in Texas and North Carolina—demonstrated high levels of mastery on the content standards established by their respective states.

The data presented in this report support the conclusion that the three new schools have been able to replicate the academic success of the two original KIPP schools. It is recommended that longitudinal studies be continued to further document the effectiveness of the KIPP educational program.

Research Design and Analysis

Each KIPP school examined in this evaluation opened in the 2001-2002 school year with students enrolled in Grade 5 only. These schools include KIPP DC/KEY Academy in the District of Columbia, Gaston College Preparatory in Gaston, North Carolina, and 3D Academy in Houston, Texas.

Three overarching questions guided the research design and analysis:

- What percentage of students is making normal educational growth each year?

- Have KIPP students made statistically significant achievement gains as compared with gains made prior to enrollment?

- Have KIPP students outperformed their traditional school counterparts?

Test-score data—including Normal Curve Equivalent (NCE) scores, scale scores, and performance levels—were collected from the schools for each student.[1] Prior achievement data were also collected for each student. Longitudinal multivariate statistical models were used to assess the effects of the KIPP educational program and its impact on student achievement.

---

[1] Appendix A in the technical report provides a detailed description of the different tests and scores used for analysis.

Gains in student achievement, as measured by the test-score data, were computed for students prior to and after attending the KIPP schools. The statistical models assessed whether KIPP students were making statistically significant gains in achievement and whether those gains were larger than those observed prior to their enrollment in the KIPP schools.[2]

In addition to the test-score data, demographic data for each student also were provided. The test-score data were disaggregated by demographic subgroup to further determine the extent to which the subgroups were making gains.

S*u*mmary of KEY Findings

KIPP DC/KEY Academy, District of Columbia

- The KIPP Grade 5 gains in math and reading were larger than those reported for any other middle/junior high school within DCPS.[3]

- All students in all subgroups increased levels of academic performance on the Stanford-9 when compared to their pretest scores. On average, KIPP students increased 23.5 NCE points in math and 12.1 NCE points in reading from fall to spring.

- A similar pattern of increased achievement was observed on the Terra Nova, suggesting that the test gains were not test-specific.

- The statistical analysis revealed that students had significantly increased reading and math performance from pretest to posttest.

- The mean gains observed for the students at KIPP DC/KEY Academy were larger than the gains obtained for the national norm.

Gaston College Preparatory School, North Carolina

- Gaston College Preparatory School showed an increase of 36 percentage points in the pass rate in reading in 2002 on the End-of-Year exam. The school had a 93 percent passing rate in reading in 2002 on the End- Of-Year exam. Only 57 percent of these same students passed the state reading test the year before while attending other schools.

- Eighty-two percent of special education students passed the state reading test in Grade 5. Only 11 percent of the special education students passed the state reading test in Grade 4 while attending other schools

- The KIPP reading achievement gains were statistically significant, whereas the prior non-KIPP gains were not.

---

[2] Only fall pretest and spring posttest data were available for the analysis in the District of Columbia. Therefore KIPP gains could not be compared to prior achievement.
[3] Data available at http://www.k12.dc.us

- The KIPP math achievement gains were statistically significant, as were the prior achievement data, showing that students continued to perform at very high levels in math while at KIPP.

- An increase in the percentage of students passing the reading End-of-Year exam was observed for all subgroups of students in the school.

- On average, 90 percent of KIPP students passed the End-Of-Year exam in math. This is an increase of 9 percent over the prior non-KIPP passing rate for the same students. An increase in passing rates was observed for all subgroups, with a slight decline for special education students.

- Gaston's Performance Composite was higher than any other school in Northampton County.[4]

3D Academy, Houston

- The average 3D Texas Assessment of Academic Skills (TAAS) passing rate for reading and math was higher than the average Houston Independent School District passing rate[5].

- On average, 98 percent of KIPP students passed the TAAS in math compared to an 89 percent passing rate for Grade 5 students in Houston Independent School District (HISD).

- On average, 88 percent of KIPP students passed the TAAS in reading compared to an 84 percent passing rate for Grade 5 students in HISD.

- Passing rates improved for all demographic subgroups of students for both the reading and math portions of the TAAS test.

- The average KIPP Stanford-9 score in math exceeded the HISD average for Grade 5 students.

- KIPP scores in reading and language were similar to HISD in language and reading as measured by the Stanford 9.

- The KIPP achievement gains for all Stanford 9 subtests were statistically significant as were the gains prior to attending KIPP. This finding suggests that students' achievement continued to increase at significant levels while at KIPP as they had prior to attending KIPP.

---

[4] Data available at http://www.dpi.state.nc.us/.
[5] Data available at http://www.houstonisd.org/.

INTRODUCTION

This report, prepared by the Education Performance Network (EPN) of New American Schools (NAS), presents the findings of an independent evaluation of three KIPP charter schools. The primary purpose of the analysis was to determine whether three new KIPP schools, in their first year of operation, were able to replicate the academic success of the two original KIPP schools in Houston and New York as measured by gains in academic achievement test scores.

Each KIPP school in this analysis opened in the 2001/2002 school year with students in Grade 5 only. These "first-year" schools include the KEY Academy in the District of Columbia, Gaston College Preparatory in North Carolina, and 3D Academy in Texas.

The following research questions were posed to guide the analyses performed:

- What percent of KIPP students are making normal educational growth each year?

- Have KIPP students made statistically significant gains when compared to prior achievement?

- Have KIPP students outperformed their traditional school counterparts?

Individual student test-score data were collected by KIPP personnel and provided to EPN to evaluate the effectiveness of the KIPP educational program in terms of its impact on student learning.

RESEARCH DESIGN/DATA ANALYSIS

Overview

The evaluation was designed to measure the academic progress that individual students made each year. Because longitudinal methods provide a richer and more accurate portrait of school effectiveness than conventional, cross-sectional approaches—which focus on different students at different times—this study examines individual student progress over time to evaluate gains in achievement and to assess the impact of the KIPP program.

The research design is a completely within-subjects interrupted time-series analysis. The gains in achievement attained after the intervention (i.e., attending the KIPP schools) were compared to the gains attained prior to attending KIPP. In this design, gains attained after the intervention can be attributed to the KIPP academic program.[6] That is, the gains attained prior to attending KIPP were compared to the gains attained after attending KIPP for one year for Gaston and 3D.

For each student at each school, individual test-score data were collected while attending Grade 5 at the KIPP schools. Prior achievement data in Grades 3 and 4 were also collected when available. However, not all KIPP schools administer the same assessments. Table 1 presents the

---

[6] The DC school was a simple pre/post design using fall-to-spring comparisons. No prior achievement data were available for this analysis.

Education Performance Network KIPP Evaluation

different tests administered at each school. (Appendix A presents a more detailed overview describing all subtests and scores used in the analyses.)

Table 1. Overview of different tests administered at each school site.

	Grade 3 (Pre-KIPP)		Grade 4 (Pre-KIPP)		Grade 5 (KIPP)	
	Stanford 9	State CRT	Stanford 9	State CRT	Stanford 9	State CRT
KIPP DC	None	None	X (proxy)[7]	None	X	None
Gaston	None	X	None	X	None	X
3D	X	X	X	X	X	X

Evaluation Approach

Assessing Normal Educational Growth

Normal Curve Equivalent (NCE) scores range from 1 to 99 and can be used to describe a student's position in the distribution of scores. NCE scores have an advantage over percentile ranks in that they approximate equal units on a scale and can be used in mathematical calculations, whereas percentile ranks cannot.

NCE scores were used to determine the extent to which students made normal educational growth each year. Students were judged to have made normal growth if they were able to maintain or exceed their position in the norm population over two consecutive years. For example, a student with an NCE reading score of 72 in Grade 3 and a score of 72 in Grade 4 would be deemed to have made normal growth. Computing the difference between NCE scores for this student would result in a change score of "0". Under this definition of normal educational growth, an NCE gain score of zero represents "normal growth," not "no growth."

In addition to the assessment of normal growth, two other analyses are presented using NCE scores. First, the *mean* NCE score describes the average performance of the group for each subtest. The mean score does not represent change over time, but rather, simply describes average performance.

Additionally, the mean NCE *gain* score is presented. The mean NCE gain score represents the average gain of the group—in this case, the school—for each subtest. An average gain of zero or greater suggests the school made normal growth, while gains less than zero suggest the school did not make normal growth.

In addition to the mean NCE gain, the percentage of students making normal growth in each school was reported for schools participating in standardized assessments. This describes the percentage of students in the school with NCE gain scores at or above zero.

Assessing Statistically Significant Gains

Although NCE scores provide the basis for the analytical approach described above, another score is needed that refers directly to content attainment. In this respect, scale scores are more suited for assessing a student's learning over time. Although scale scores can be used to

---

[7] The Intermediate 1 Stanford 9 was administered in the fall of 2001 (when the students were in Grade 5) and serves as a proxy for prior Grade 4 achievement.

compare a student's progress over time, they cannot be used to compare one subtest to another. Thus, in the case of the Stanford 9—with scale scores ranging from 1 to 999—a scale score of 500 in reading is not comparable to a 500 in math.

Scale scores can be used to measure different forms and levels of the Stanford 9 test. For example, if a student had a scale score of 500 in math in Grade 4 and in Grade 5 had a scale score of 500 in math, this student is deemed to have made no growth.

To assess the extent to which the school made statistically significant gains over time, multivariate statistical analyses were performed using scale scores for available subtests as the dependent variables. All statistical analyses were completely within-subjects, as no comparison group data were available for analysis. Each statistical test performed is described under the school subheadings in the next section.

In order to assess the effects on the dependent variables separately, a repeated contrast test was also performed to determine the statistical significance of the individual dependent variables. Using these comparisons, we test whether students made statistically significant gains[8] on each subtest by comparing their performance from Grades 3 to 4 and from Grades 4 to 5.

The repeated contrast was used to support an *a priori* hypothesis that these students were not making statistically significant gains prior to attending KIPP. However, it was expected that these same students would make statistically significant gains following their attendance at KIPP. Findings in support of this hypothesis would provide positive evidence that the KIPP instructional program was more successful than prior schooling for these same students.

Comparing KIPP Students to Traditional Students

Individual student-level data from other schools were not provided and thus, could not serve as a basis for comparing the performance of KIPP students with that of their traditional school counterparts. Therefore, two proxy measures were used to assess the relative performance of students enrolled in the KIPP schools. First, district-level aggregate achievement data were used to compare the absolute levels of performance of traditional and KIPP schools. For example, the district average passing rate[9] in reading was compared to the KIPP average passing rate in reading at the Houston school.

In addition, the Technical Data Report for the Stanford 9 was used to compute gain scores for the norm group[10]. These gain scores were computed based on cross-sectional cohorts of student scores at the 50th percentile. The mean scale score gain for the KIPP students was then compared to the mean scale score gain computed for the norm group.

---

[8] Alpha was set at .01 for all tests with accompanying effect size measures.
[9] Passing rates are based on students with test scores, excluding those exempted from the assessment or for whom there are no data.
[10] Stanford Achievement Test Series Ninth Edition: Technical Data Report. Harcourt Educational Measurement. San Antonio, 1997.

Data and Analysis

KIPP DC/KEY Academy

Prior achievement scores (i.e., Grade 4 scores) on the Stanford 9 test were not available for students at KIPP DC/KEY Academy in the District of Columbia. Instead, pretest scores obtained from an administration of the Stanford 9 at the beginning of the school year served as a proxy measure for prior achievement.[11]

We analyze the pre/post Stanford 9 data using a doubly multivariate analysis of variance (MANOVA). Additionally, a repeated contrast was performed to test the statistical significance of each variable independently.

In addition to the Stanford 9, the KEY Academy opted to administer the Terra Nova standardized assessment. The test was administered at the beginning and end of the school year, providing pretest and posttest data for analysis. The mean NCE score for the pretest and posttest was computed using the Terra Nova data. These data served as one means of confirming the growth pattern observed on the Stanford 9.

Beyond the test-score data, student-level data on gender, Title 1 eligibility, ethnicity, and Limited English Proficiency (LEP) were provided for 80 participants. All students in the school were African American, none were LEP, 51% were female, and 80% were eligible for free and reduced lunch (a proxy measure for economic status). The participation rates are presented in Table 2.

Table 2. KIPP DC/KEY Academy SAT 9 Participation rates

		Pretest	Posttest
Math	Total number of eligible students	80	80
	Total number of tested students	80	80
	Total number exempt/missing data	0	0
	Percent of population exempt/missing data	0%	0%
	Percent of population tested	100%	100%
Reading	Total number of eligible students	80	80
	Total number of tested students	79	80
	Total number exempt/missing data	1[12]	0
	Percent of population exempt/missing data	1%	0%
	Percent of population tested	99%	100%

Gaston College Preparatory

The data from Gaston College Preparatory in North Carolina included performance levels and scale scores from the State End-of-Year exams in reading and math. Prior achievement scores on the End-of-Year exams, including Grade 3 and 4 reading and math scores, were collected for all students.

---

[11] The scores obtained from the pretest are from the Intermediate 1 Spring form of the Stanford 9.
[12] One student had partial reading data, therefore a total reading score was not reported.

The End-of -Year exam data are reported in three ways. The tests have been vertically equated with scale scores representing a developmental continuum permitting the use of longitudinal statistical analyses. The scale scores for reading range from 114 to 182 in Grades 3 through 5, and the math scale scores range from 218 through 295 in Grades 3 through 5.

Each student was assigned a performance standard—one, two, three, or four—based on his or her scale score.[13] A performance standard of "one" represents the lowest level of performance, while a "four" corresponds to the highest performance level. Students who performed at levels three or four are deemed to have passed the test. A dichotomous variable was created for analysis where "1" (i.e., performance standard three or four) indicates a student passed the test and "0" (i.e., performance standard one or two) indicates that the student failed.

Scale scores were used for a completely within- subjects multivariate analysis of variance, where the dependent variables were reading and math scale scores. A repeated contrast was performed with the *a priori* hypothesis that students did not make statistically significant gains in academic achievement from Grade 3 to 4, but did make significant gains from Grade 4 to 5.

The proportion of students passing the End-of-Year exam was computed for each of the three years. The data were then disaggregated by subgroups, as defined by the demographic data provided. These data were plotted as a time-series to examine the increase in passing rates over time.

Demographic data for 73 students at Gaston School—including gender, ethnicity, LEP, special education status, and Title 1 eligibility—also were provided. Ninety-five percent of the students were classified as minority (all but two were identified as African-American), 15% were receiving special education services, 82% were eligible for free or reduced-price lunch, and 47% were female. The participation rates are presented in Table 3.

Table 3. Gaston End-of-Year Exam participation rates

		Grade 3	Grade 4	Grade 5
Math	Total number of eligible students	73	73	73
	Total number of tested students	60	69	73
	Total number exempt/missing data	13	4	0
	Percent of population exempt/missing data	18%	5%	0%
	Percent of population tested	82%	95%	100%
Reading	Total number of eligible students	73	73	73
	Total number of tested students	55	67	73
	Total number exempt/missing data	18	6	0
	Percent of population exempt/missing data	25%	8%	0%
	Percent of population tested	75%	92%	100%

3D Academy

Data for 3D Academy in Houston included the Texas Assessment of Academic Skills (TAAS) and Stanford 9 test-score data for each student over a three-year period. The TAAS data included the Texas Learning Index (TLI) scores as well as a dichotomous variable coded "1" for students

---

[13] The State of North Carolina used the Contrasting Groups method to create Performance Standards.

who passed the TAAS test and "0" for those who did not. The TLI scores have not been placed on a developmental continuum and cannot be used to measure progress over time. Therefore, the percentage of students passing the TAAS test was computed for each year and for each demographic subgroup to measure incremental progress over time.

As in the case of the KEY Academy, Stanford 9 scores—including NCE and scale scores— were collected for 3D Academy students. Therefore, the same NCE analyses were performed for 3D and DC.[14]

A completely within-subjects doubly multivariate analysis of variance was performed using Stanford 9 scale scores. Total reading, total math, and total language scores served as the dependent variables. Similar to the analysis performed for the Gaston School, an *a priori* hypothesis was posed indicating that students did not make statistically significant gains in achievement from Grade 3 to Grade 4, but did make statistically significant gains in achievement from Grade 4 to Grade 5.

Demographic data were collected for 70 enrolled students at 3D Academy. Forty-two percent were female, 27.3% African American, 68.2% Hispanic, 4.5% White, 61% LEP, 10% were receiving special education, and 86% were eligible for free or reduced-price lunch. Stanford 9 and TAAS participation rates are presented in Tables 4 and 5.

Table 4. 3D Stanford 9 participation rates

		Grade 3	Grade 4	Grade 5
Math	Total number of eligible students	70	70	70
	Total number of tested students	47	54	70
	Total number exempt/missing data	23	16	0
	Percent of population exempt/missing data	33%	23%	0%
	Percent of population tested	67%	77%	100%
Reading	Total number of eligible students	70	70	70
	Total number of tested students	47	54	70
	Total number exempt/missing data	23	16	0
	Percent of population exempt/missing data	33%	23%	0%
	Percent of population tested	67%	77%	100%
Language	Total number of eligible students	70	70	70
	Total number of tested students	47	53	70
	Total number exempt/missing data	23	17	0
	Percent of population exempt/missing data	33%	24%	0%
	Percent of population tested	67%	76%	100%

---

[14] The language portion of the Stanford 9 is administered at Houston and not in DC.

Table 5. 3D TAAS participation rates

		Grade 3	Grade 4	Grade 5
Math	Total number of eligible students	70	70	70
	Total number of tested students	48	55	66
	Total number exempt/missing data	22	15	4
	Percent of population exempt/missing data	31%	21%	6%
	Percent of population tested	69%	79%	94%
Reading	Total number of eligible students	70	70	70
	Total number of tested students	46	54	64
	Total number exempt/missing data	24	16	6
	Percent of population exempt/missing data	34%	23%	9%
	Percent of population tested	66%	77%	91%

RESULTS

KIPP DC/KEY Academy

Table 6 presents the mean NCE scores for all subtests, including pretest scores and posttest scores, on the Stanford 9 test. Table 7 provides overall descriptive statistics for all subtests on the Terra Nova. The data suggest a similar pattern of achievement as measured by both tests.

Table 6. Mean NCE Scores (SAT 9), KIPP DC/KEY Academy

	Pretest	Posttest	Difference
Total Math	41.19	64.73	23.54
Problem Solving	40.95	62.56	21.61
Procedures	43.84	65.74	21.90
Total Reading	34.09	46.22	12.13
Reading Vocabulary	34.47	46.76	12.29
Reading Comprehension	36.66	46.83	10.17

Table 7. Mean NCE Scores (Terra Nova), KEY Academy

	Pretest	Posttest	Difference
Total Math	33.48	52.78	19.3
Math Computation	47.61	77.21	29.6
Math Composite	38.48	65.95	27.47
Total Reading	34.37	49.56	15.19
Reading Vocabulary	43.11	46.12	3.01
Reading Comprehension	38.35	47.73	9.38

The results from KIPP DC/KEY Academy in the District of Columbia demonstrated very large NCE gains as measured by the Stanford 9. These gains are based on the difference between the school-wide average pretest from the school-wide average posttest. The mean reading gain was 12.13 NCEs, while the mean math gain was 23.54 NCEs. Although the gains were measured from a fall-to-spring administration of the test rather than a spring-to-spring administration, the NCE gains are substantial.

Education Performance Network KIPP Evaluation

The mean NCE gain scores are reported in Figure 1. These gains are based on students with matched test scores from pretest to posttest. The changes are much larger than expected given each student's pretest score information. The figure further illustrates that, although reading gains were large, math gains were almost twice the size of those for reading.

Figure 1. Mean SAT 9 NCE Gain Score by Subtest, KEY Academy

	Total Math	Problem Solving	Procedures	Total Reading	Reading Vocabulary	Reading Comprehension
KIPP Mean	23.72	21.74	22.13	12.29	12.33	10.27
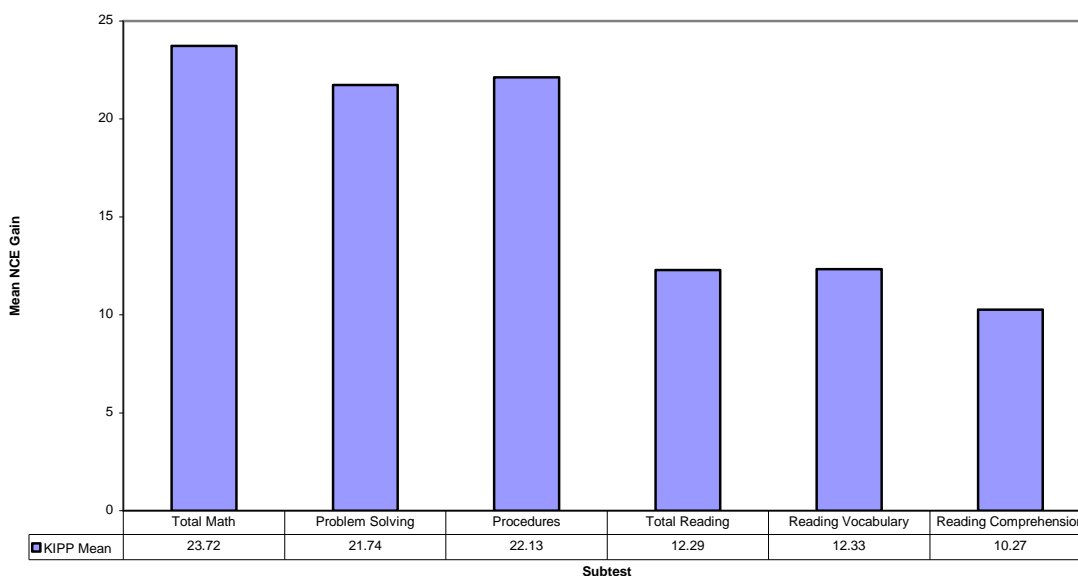
Table 8 presents the percentage of students with NCE gains of zero or higher from the pretest to the posttest. The data reveal that 97% of students in math and 93% in reading showed NCE gains at or above zero. The gains for the Title 1 subgroup, traditionally defined as low performing, also made very large achievement gains.

Table 8. Percent Making Normal Educational Growth (SAT 9), KEY Academy

	Total	Male	Female	Title 1
Total Math	**97%**	**95%**	**100%**	**98%**
Problem Solving	95%	92%	97%	97%
Procedures	93%	95%	92%	93%
Total Reading	**93%**	**87%**	**100%**	**95%**
Reading Vocabulary	89%	92%	87%	90%
Reading Comprehension	87%	80%	94%	86%

The District of Columbia Public School District (DCPS) reports the school-wide aggregate NCE for each school in its system[15]. Therefore, a direct Grade 5 comparison cannot be assessed. However, the school-wide NCE gain score over subsequent school years, as computed by DCPS, can serve as a proxy measure. Using this method of comparison, the observed KIPP gains in reading and math were larger than gains reported for any other middle/junior high school within DCPS in 2002.

---

[15] Data available at http://www.k12.dc.us/.

Education Performance Network KIPP Evaluation

The results of the statistical gain score analysis, where reading and math scale scores were the dependent variables for a doubly multivariate analysis of variance, confirmed the NCE gain score pattern, indicating that statistically significant academic gains were attained from pretest to posttest. The value of Wilks' Lambda = .143, where the omnibus multivariate $F (2, 73) = 218.62$, $p < .01$.

Table 9 presents the means and standard deviations, and Table 10 presents the results of the MANOVA and repeated contrasts performed to examine each dependent variable separately. The results presented in Table 10 indicate that scores on the posttest for both reading and math were significantly higher than on the pretest.

This statistically significant finding suggests that the gains in Grade 5 test scores did not occur by chance. Although the gains are very large, pre/post designs limit causal inferences. This finding warrants further study, as this analysis did not examine the growth of these students prior to entering KIPP, nor did it include the use of a comparison group. Nevertheless, these findings provide very positive preliminary evidence to support the effectiveness of the KIPP program in DC.

Table 9. SAT 9 Means and Standard Deviations for Reading and Math, KEY Academy

	Mean	Standard Deviation	N
Reading Pretest	617.63	33.48	75
Reading Posttest	647.32	28.52	75
Math Pretest	620.52	32.82	75
Math Posttest	673.0	27.66	75

Table 10. MANOVA results for Reading and Math scores, KEY Academy

	SS	df	MS	F	η^2
Math	33063.527	1	33063.527	191.914*	.72
Reading	103280.64	1	103280.64	403.198*	.85
Error (Read)	12748.973	74	172.283		
Error (Math)	18955.36	74	256.154		

*$p < .01$

Figure 2 reports the mean scale score gain for students at KEY Academy from the fall pretest to the spring posttest administration and compares that to the mean scale score gain for the norm population from spring to spring as described in the SAT 9 technical manual. As revealed in the data provided, the KIPP achievement levels were approximately 2.3 times larger in total math and 2 times larger in total reading than those observed by students participating in the norm group.[16]

---

[16] The gains presented in Figure 2 differ slightly from those reported in Table 9 as cases were deleted listwise for the MANOVA and no cases were deleted for this analysis.

Education Performance Network KIPP Evaluation

Figure 2. Mean SAT 9 Scale Score Gain Comparison, DC

	Total Math	Problem Solving	Procedures	Total Reading	Reading Vocabulary	Reading Comprehension
☐ KIPP Gain	51.85	41.14	71.54	29.46	32.15	28.66
■ Norm Gain	22.6	19	28.3	14.9	17.5	13.4

Subtests
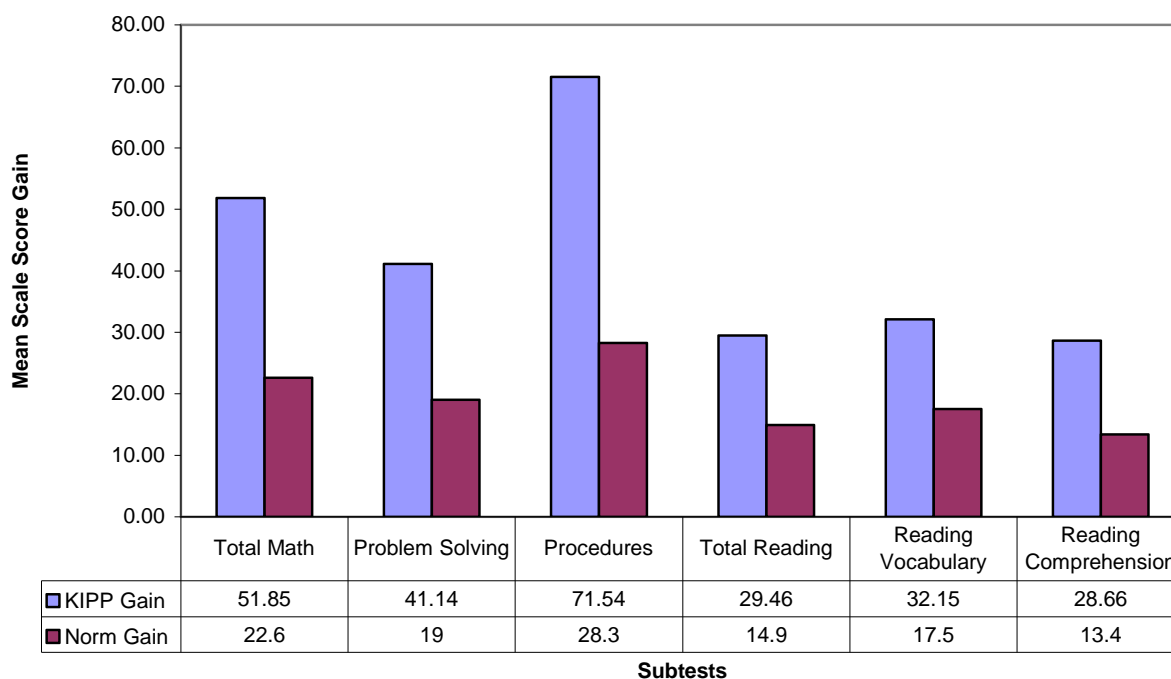
*The norm group gains were computed using spring to spring comparisons of cross-sectional cohorts at the 50th percentile

The findings presented in this section provide preliminary positive evidence that the KIPP instructional program in DC has led to a large increase in achievement gains beyond what could have occurred by chance alone. Additionally, the gains appear to exceed the observed cross-sectional gains in the norm group as presented in Figure 2.

Gaston College Preparatory

Results from the Gaston school provide evidence to suggest that, relative to prior achievement, students significantly improved their performance as measured by the state-designed End-of-Year exams.

The data in Tables 11 and 12 represent the proportion of students passing the State test in Grades 3 through 5 in reading and math, respectively[17]. The scores for Grades 3 and 4 were attained prior to attending KIPP. Figures 3 and 4 provide visual representation of the same data, supporting the hypothesis that levels of academic achievement increased for all students after only one year of schooling at the Gaston School.

As shown in Figure 3, the increase in reading achievement in the year after students enrolled in the Gaston school represents a reversal of a downward trend evidenced in the year prior to their enrollment. Figure 4 shows that although the absolute level of performance for math is higher than in the previous year, examination of the growth slope suggests a slight *relative* decline. This may be the result of a ceiling effect. That is, students had already progressed on the scale substantially and had less room to grow.

---

[17] Appendix D presents an alternative calculation of the passing rates where all students, including those classified as exempt or with missing data, are included in the calculation.

Education Performance Network KIPP Evaluation

The data for reading and math suggest two specific accomplishments. First, reading achievement improved over time across all demographic subgroups. This pattern also was observed across all subgroups in math, except for white children (who maintained high levels of performance) and special education students (who showed a slight decline in achievement). Second, Gaston students outperformed those enrolled in all other schools in Northhampton County[18].

Table 11. Percent of Students at Proficient or Above (Reading), Gaston

	Total	Male	Female	White	Minority	Title 1	SPED
Grade 3 (Pre-KIPP)	67%	59%	77%	100%	65%	65%	44%
Grade 4 (Pre-KIPP)	57%	46%	69%	100%	54%	54%	11%
Grade 5 (KIPP)	93%	92%	94%	100%	93%	92%	82%

Table 12. Percent of Students at Proficient or Above (Math), Gaston

	Total	Male	Female	White	Minority	Title 1	SPED
Grade 3 (Pre-KIPP)	64%	65%	63%	100%	62%	58%	38%
Grade 4 (Pre-KIPP)	81%	78%	83%	100%	79%	80%	67%
Grade 5 (KIPP)	90%	90%	91%	100%	90%	90%	64%

Figure 3. Percent of Students at or Above Proficient in Reading, Gaston

Total

	Grade 3 (Pre-KIPP)	Grade 4 (Pre-KIPP)	Grade 5 (KIPP)
Total	67%	57%	93%

---

[18] The data available at http://www.dpi.state.nc.us/.

Education Performance Network KIPP Evaluation
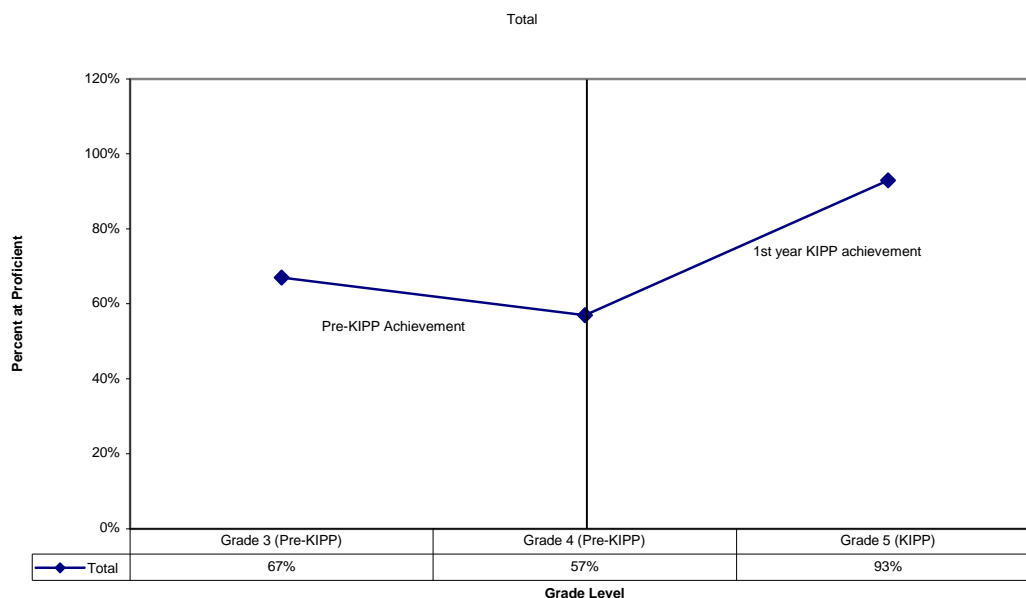
Figure 4. Percent of Students at or Above Proficient in Math, Gaston

	Grade 3 (Pre-KIPP)	Grade 4 (Pre-KIPP)	Grade 5 (KIPP)
Total	64%	81%	90%

Grade Level
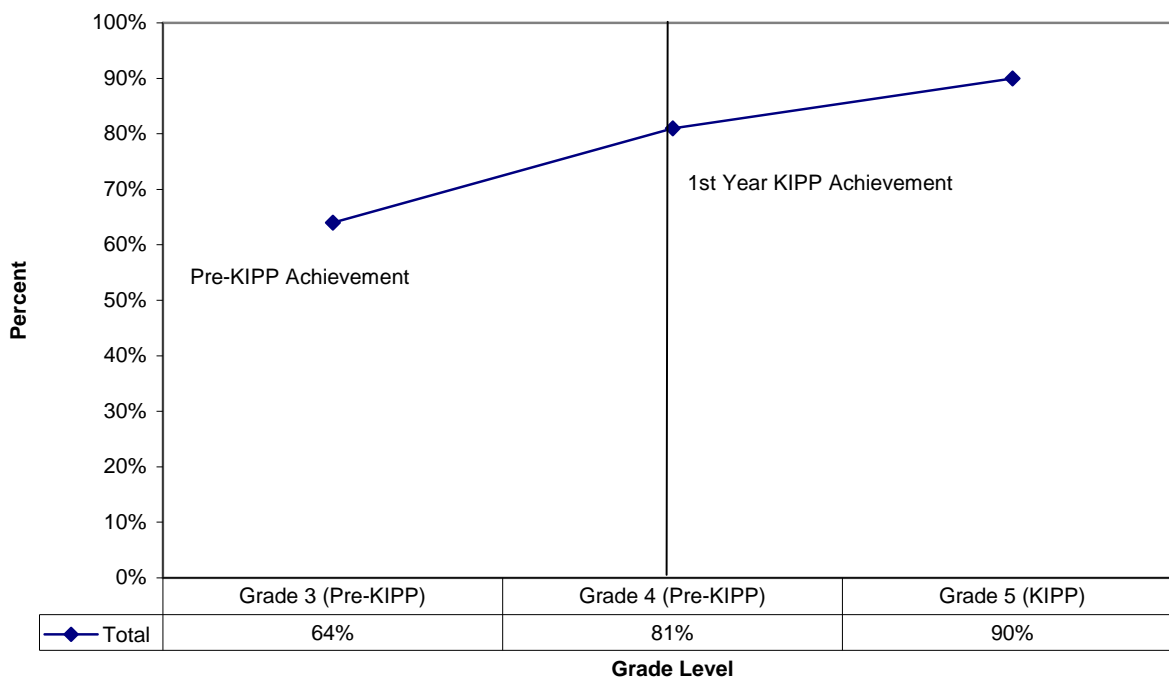
To assess the significance of the gains for reading and math a doubly multivariate analysis of variance was performed, employing reading and math scale scores for the dependent variables. A repeated contrast was also performed to compare Grade 3 reading and math with Grade 4, and Grade 4 reading and math with Grade 5. It was expected that the pre-KIPP gains would not be statistically significant, while the KIPP gains would be. Partial η^2 was computed for each significant contrast to measure the strength of association.

The value of Wilks' Lambda for the omnibus multivariate test was .082, $p < .01$. The results of the repeated contrast revealed that the Grade 4 reading scores were not significantly higher than the Grade 3 reading scores, $F(1,50) = 6.77$ $p > .01$. However, the Grade 5 reading scores were significantly higher than the Grade 4 reading scores, $F(1, 50) = 235.91$, $p < .01$, $\eta^2 = .83$. The Grade 4 math scores were significantly higher than the Grade 3 math scores, $F(1, 50) = 35.99$, $p < .01$, $\eta^2 = .42$, and the Grade 5 math scores were also significantly higher than the Grade 4 math scores, $F(1, 50) = 207.33$, $p < .01$, $\eta^2 = .81$. Table 13 presents the means and standard deviations of the dependent variables used in the analysis. Appendix C presents the mean scale scores by demographic subgroup. [19]

These findings provide evidence to partially support the hypothesis posed. The non-significant gains in reading from Grades 3 to 4 suggest that these students were not achieving at very high levels prior to attending KIPP. However, after attending KIPP in Grade 5, these same students achieved impressive academic gains on the state test. This finding provides positive support for the effectiveness of the academic reading program at the Gaston School.

---

[19] The scale score data differ in Appendix C slightly from those presented in Table 13 as listwise deletion was used for the analyses and no scores were deleted from this analysis.

Education Performance Network KIPP Evaluation

The math gains prior to attending KIPP were statistically significant, as were the gains occurring at KIPP. Although these students were performing at high levels in math prior to attending KIPP, the effect size measure (η^2) is noticeably larger for the KIPP gains than for the prior achievement gains. This finding suggests that the achievement gains at KIPP were more pronounced than achievement gains prior to attending KIPP.

Table 13. Means and Standard Deviations for EOY Reading and Math Exam, Gaston

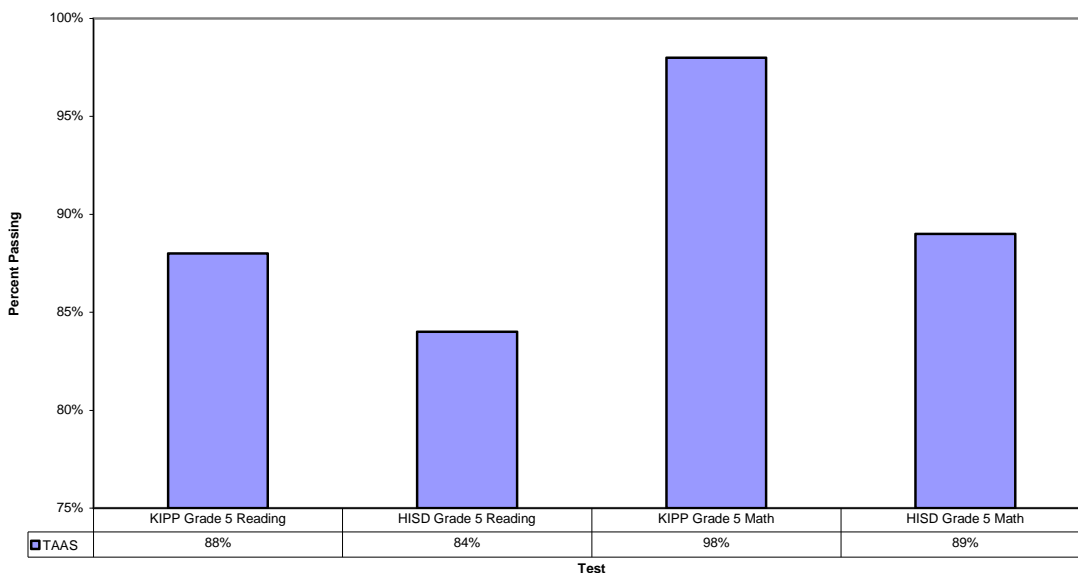|	Mean	Standard Deviation	N
Reading Grade 3	145.29	8.52	51
Reading Grade 4	147.41	7.99	51
Reading Grade 5	157.86	6.02	51
Math Grade 3	250.31	7.2	51
Math Grade 4	254.8	8.7	51
Math Grade 5	261.86	8.31	51

The results of the analyses in this section provide positive evidence that KIPP students in North Carolina have made statistically significant gains in reading and math when compared to prior levels of achievement. On average, KIPP students have demonstrated high levels of mastery on the state exams and outperformed all other county schools.

3D Academy

Figure 5 provides results from the 2002 administration of the Texas TAAS test for students enrolled in 3D Academy and for those enrolled in traditional schools in the Houston Independent School District (HISD). The data describe the percentage of KIPP students passing the TAAS in reading and math as compared to the percentage of HISD students passing.[20] As indicated by the data, KIPP students had a higher passing rate in 2002 than the average HISD student. Even more noteworthy is the 98% passing rate in math for the students enrolled at 3D Academy.

---

[20] The data reported were collected from the HISD web site for Grade 5 students in 2002.

Education Performance Network KIPP Evaluation

Figure 5. 3D/HISD Grade 5 TAAS Comparison

	KIPP Grade 5 Reading	HISD Grade 5 Reading	KIPP Grade 5 Math	HISD Grade 5 Math
□ TAAS	88%	84%	98%	89%

The data in Figure 5 above only provide a cross-section of test scores for one year. To further examine the TAAS results at 3D, individual student passing rates were collected for two years prior to attending KIPP. The data, disaggregated by demographic subgroup for both reading and math, are displayed in Figures 6 and 7, respectively. The disaggregated longitudinal data provide evidence to suggest that all demographic subgroups made improvements in math when compared to Grade 4 data. Additionally, most subgroups increased in reading when compared to Grade 4 percentages.

Figure 6. TAAS Reading by Subgroup, 3D
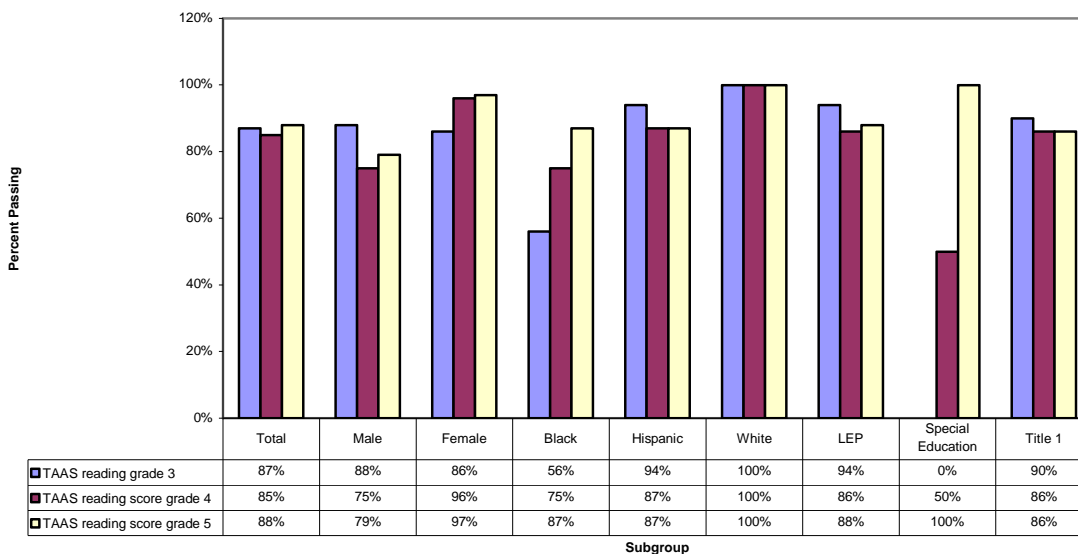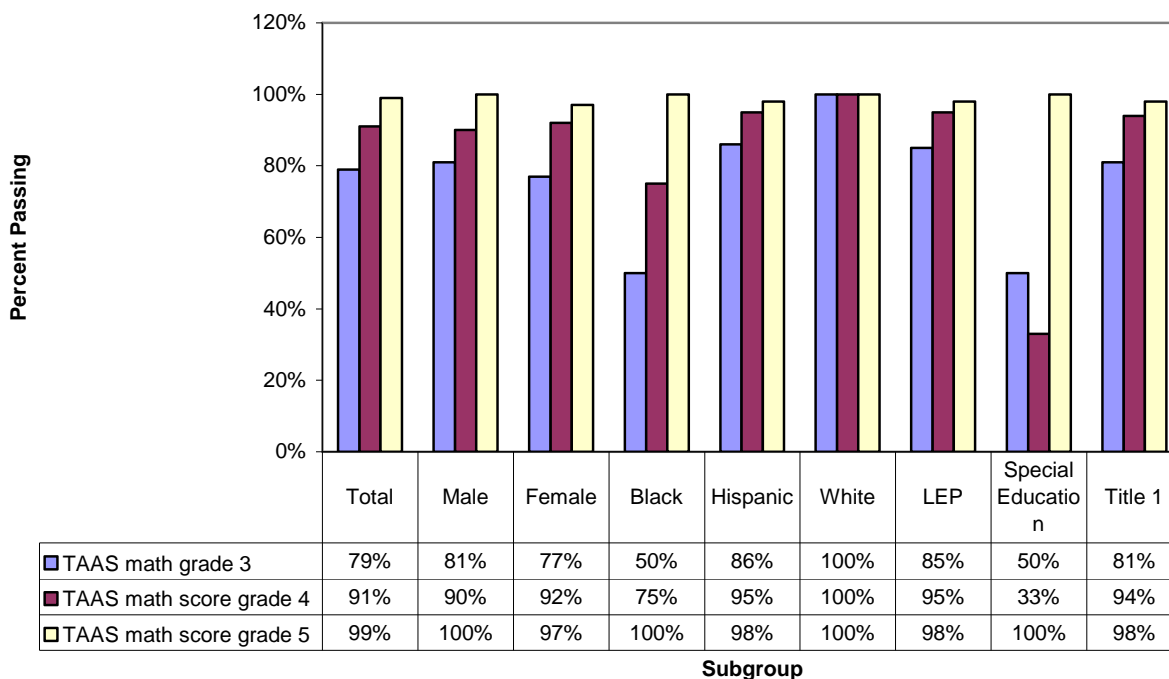
	Total	Male	Female	Black	Hispanic	White	LEP	Special Education	Title 1
□ TAAS reading grade 3	87%	88%	86%	56%	94%	100%	94%	0%	90%
■ TAAS reading score grade 4	85%	75%	96%	75%	87%	100%	86%	50%	86%
□ TAAS reading score grade 5	88%	79%	97%	87%	87%	100%	88%	100%	86%

Education Performance Network KIPP Evaluation
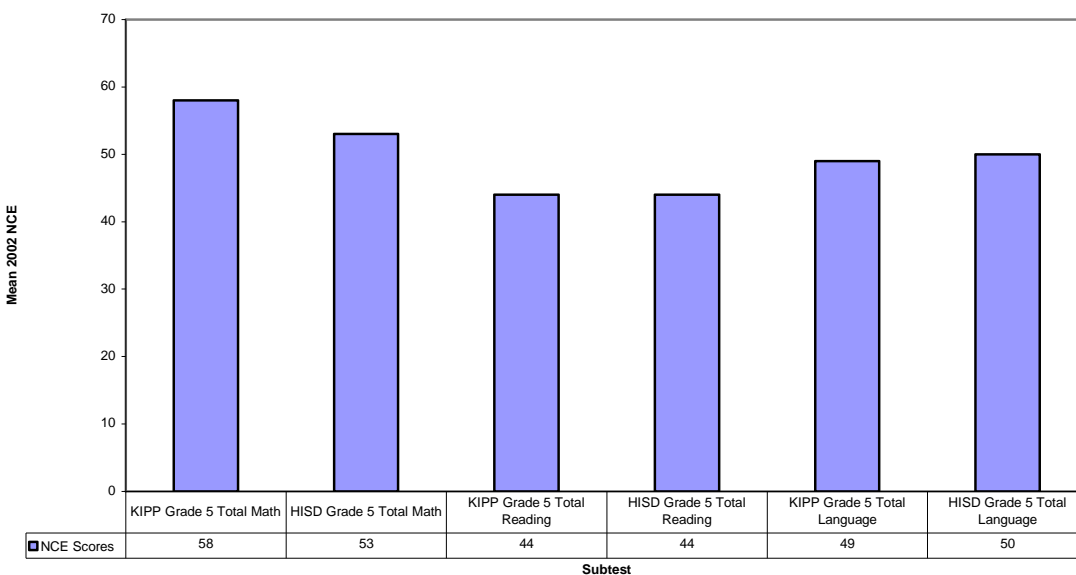
All Grade 5 students at 3D participated in the Stanford 9 English version of the assessment. Prior to attending KIPP, some students were administered the Stanford 9 Spanish version. Students participating in the Spanish version of the Stanford 9 in either Grade 3 or 4 were excluded from the gain score analysis. The excluded students comprised 13% in Grade 3 and approximately 1.5% in Grade 4.

The data in Table 14 suggest that more Grade 5 students made normal growth in reading than in the previous non-KIPP year. Additionally, the percentage of students making normal educational progress in Grade 5 math was substantially larger than had been previously attained. The language scores show a decline in the percentage of students making normal educational growth as compared to student performance the previous school year.

Table 14. Percent Making Normal Educational Growth (SAT 9), 3D

	Grade 3 to 4	**Grade 4 to 5**
Total Reading	43.6%	52%
Total Math	35.9%	62%
Total Language	61.5%	52%

Table 15 reports the mean NCE gain scores by subtest over three consecutive years. The data indicate that—with the exception of math procedures and total language—negative change scores occurred across all subtests for students prior to attending KIPP. Conversely, positive change scores occurred for almost all subtests the following year, the first year at KIPP, with the exception of language expression and total language.

Table 15. SAT 9 NCE gain scores, 3D

	Grade 3 to 4 (Pre-KIPP)	Grade 4 to 5 (KIPP)
Total reading	-5.3	1.08
Vocabulary	-9.09	3.6
Reading comprehension	-2.28	0.32
Total math	-0.85	2.10
Problem solving	-3.28	1.49
Procedures	1.47	1.13
Total Language	1.94	-0.84
Language Mechanics		1.2
Language Expression		-2.3

HISD does not report the student-level NCE gains from spring to spring. Therefore, the observed NCE gains at 3D cannot be compared to student-level NCE gains in Houston. However, the difference between NCE Grade 4 scores in 2001 and the Grade 5 scores in 2002 can serve as a quasi-longitudinal proxy measure for comparison[21]. To compute the HISD change scores, the mean NCE score in Grade 4 (2001) for reading, math, and language was compared to the mean NCE score in Grade 5 (2002) for each subtest. These data are presented in Table 16. The data suggest that, while HISD students showed a decline in the mean NCE score from Grade 4 to 5, KIPP students showed an increase in mean NCE gain score from Grade 4 to 5 in reading

---

[21] These data were obtained from the HISD web site at http://www.houstonisd.org/.

Education Performance Network KIPP Evaluation

and math with a slight negative mean gain score for language. Appendix A reports the mean NCE gain by defined demographic subgroup for KIPP students.

Table 16. KIPP/Houston SAT 9 NCE Gains, 3D

	HISD 4 to 5	KIPP 4 to 5
Total reading NCE gain score	-4	1.08
Total math NCE gain score	-2	2.10
Total language NCE gain score	-5	-0.84

To examine the results statistically, a completely within-subjects doubly multivariate analysis of variance was performed on the three dependent variables: total reading, total math, and total language. An *a priori* hypothesis was set predicting that gains occurring between Grade 3 and 4 would not be significant, while the gains occurring between Grade 4 and 5 would be significant. To examine this hypothesis, a repeated contrast was performed for each dependent variable.

The value of Wilks' lambda for the omnibus multivariate test was .185, $p < .01$. The results of the repeated contrast for reading revealed that the Grade 4 scores were significantly higher than the Grade 3 scores, $F (1, 26) = 31.44$, $p < .01$, $\eta^2 = .55$ and the Grade 5 scores were significantly higher than the Grade 4 scores, $F (1, 26) = 34.37$, $p < .01$, $\eta^2 = .57$. The Grade 4 math scores were significantly higher than the Grade 3 scores, $F (1, 26) = 23.82$, $p < .01$, $\eta^2 = .48$ and the Grade 5 scores were significantly higher than the Grade 4 scores, $F (1, 26) = 82.63$, $p < .01$, $\eta^2 = .76$. The Grade 4 language scores were significantly higher than the Grade 3 scores, $F (1, 26) = 35.84$, $p < .01$, $\eta^2 = .58$ and the Grade 5 language scores were significantly higher than the Grade 4 scores, $F (1, 26) = 8.18$, $p < .01$, $\eta^2 = .26$. Table 17 presents the means and standard deviations used in the analysis.

Table 17. SAT 9 Means and Standard Deviations for Stanford Reading, Math, and Language, 3D

	Mean	Standard Deviation	N
Reading Grade 3	592.89	37.04	27
Reading Grade 4	617.11	27.89	27
Reading Grade 5	639.15	25.93	27
Math Grade 3	600.15	42.17	27
Math Grade 4	626.56	31.59	27
Math Grade 5	659.41	26.28	27
Language Grade 3	591.52	33.95	27
Language Grade 4	624.70	28.79	27
Language Grade 5	636.59	33.02	27

These findings suggest that statistically significant increases in test scores have occurred for these students in reading, math, and language, both while enrolled at 3D Academy and prior to enrollment. The *a priori* hypothesis was not supported given that these students were making significant gains on the Stanford 9 prior to attending the KIPP school. However, these same students continued to attain very high levels of achievement after attending 3D.

Figure 9 compares the scale score gains attained prior to attending 3D Academy to those for the norm group. The KIPP scale score gains were computed using spring-to-spring score reports

for individual students.[22] The data in Figure 9 indicate that students were making larger gains than the norm group prior to attending the KIPP school in all subtests except reading vocabulary.

Figure 9. Mean SAT 9 Scale Score Gain Comparison (3 to 4), 3D

	Total Math	Problem Solving	Procedures	Total Reading	Reading Vocabulary	Reading Comprehension	Total Language
☐ Pre-KIPP Gain 3 to 4	33.98	23.23	65.28	25.44	26.00	27.19	36.75
■ Norm Gain 3 to 4	24.7	15.9	37.7	23.3	30.1	19.8	18.3

Subtest
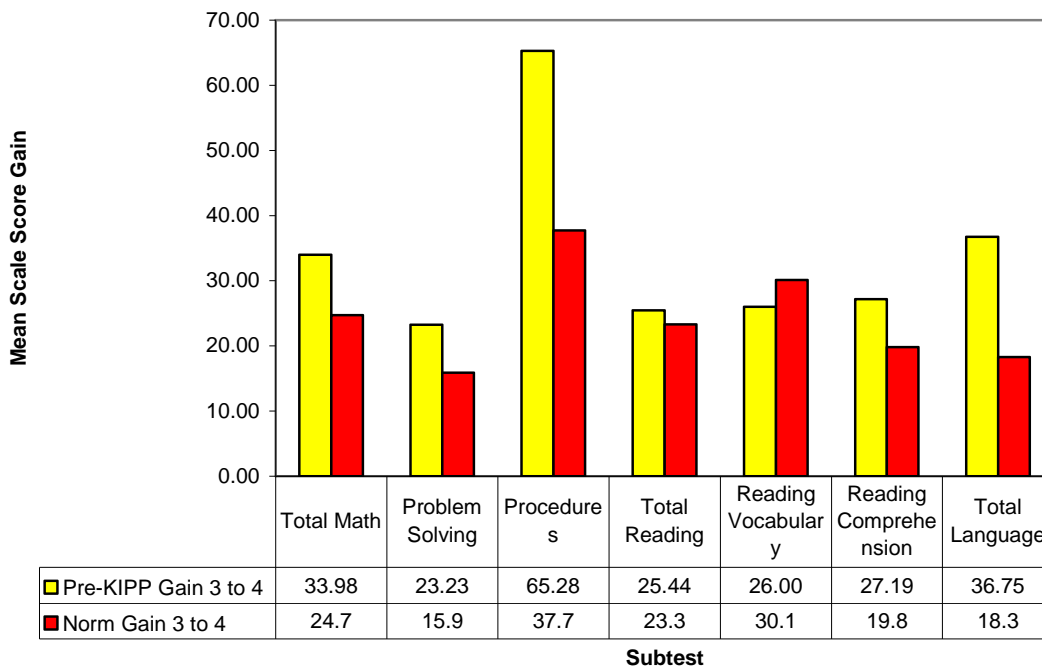
Figure 10 compares the scale score gains attained while attending 3D Academy in Grade 5 with those made by the norm group. The data illustrate the larger gains attained by KIPP students in Grade 5 as compared with gains of the norm group. KIPP gains were larger in all areas except for total language and language expression.

---

[22] The mean scale scores used for this analysis differ slightly from those presented above in Table 17 as cases were deleted listwise for the statistical analysis and only students participating in the Spanish version of the test were excluded from this analysis.

Figure 10. Mean SAT 9 Scale Score Gain Comparison (4 to 5), 3D

	Total Math	Problem Solving	Procedures	Total Reading	Reading Vocabulary	Reading Comprehension	Total Language	Language Mechanics	Language Expression
☐ KIPP Gain 4 to 5	25.33	20.18	33.67	26.56	33.29	21.67	12.16	14.85	11.64
■ Norm Gain 4 to 5	22.6	19	28.3	14.9	17.5	13.4	12.6	10.7	13.8

Subtests
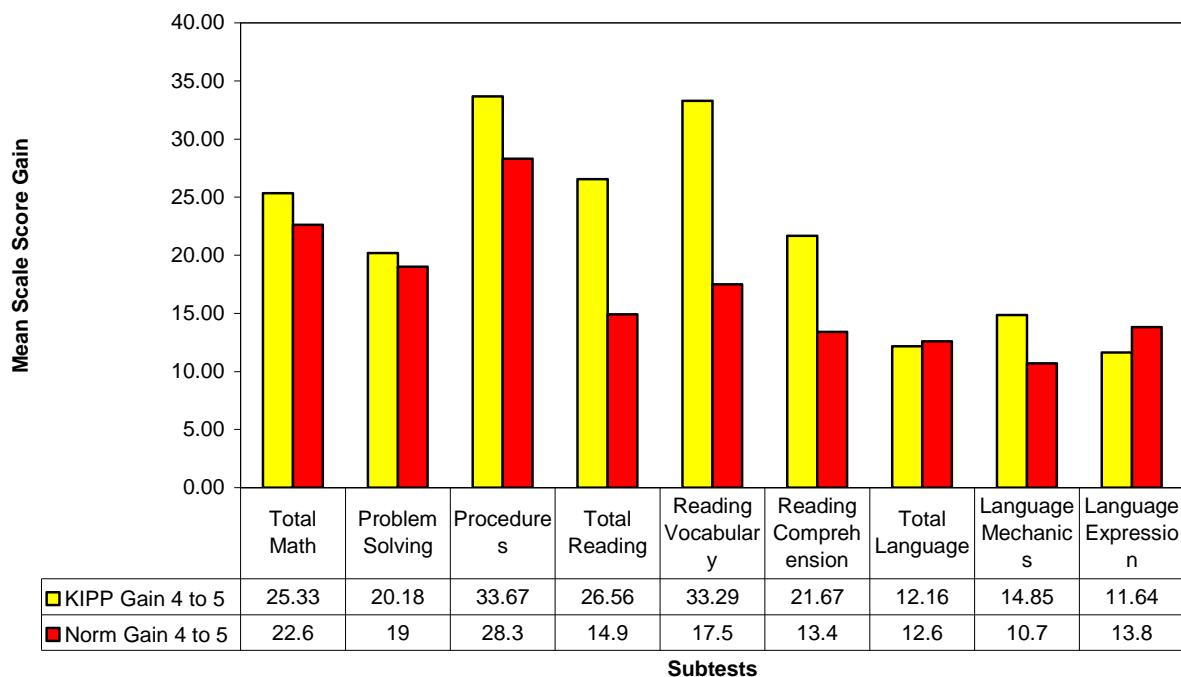
The results presented in this section suggest that students enrolled at 3D Academy had higher gains on the Stanford 9 than their counterparts at HISD schools. KIPP students exhibited significantly increased academic achievement in reading, math, and language. Further, the data suggest that KIPP students had larger academic gains than the norm group.

Although these findings suggest large gains for students while attending 3D Academy, these same students were making statistically significant gains prior to their enrollment. This suggests that the students continued to advance academically, as they had prior to attending the KIPP school.

CONCLUSIONS

The first year data from all three schools provide positive evidence to support the effectiveness of the KIPP instructional program. Observed academic gains exceed what could have occurred by chance and exceed the performance of respective district performance. Each school increased levels of academic achievement performance for students, regardless of background or label.

The gains at KIPP DC/KEY Academy are notably large and reveal that almost all students in the school have made normal educational growth and have made statistically significant gains over time. Additionally, the observed KIPP gains are larger than the gains reported for any other junior/middle school within DCPS. The addition of prior achievement data will add to the strength of this finding.

The high levels of mastery on the state criterion-referenced tests in Texas and North Carolina suggest that KIPP students are demonstrating high levels of mastery on the state defined content

standards. Additionally, these students demonstrated higher levels of performance on the state tests than did their traditional school counterparts in each respective district.

The results of the statistical analyses for Gaston and 3D are statistically significant and indicate impressive academic gains for the students enrolled. Although partial support for the *a priori* hypothesis was found at Gaston, with no support at 3D, this does not disqualify the remarkable academic achievement gains observed. Rather, it suggests that KIPP students continued to increase in achievement at KIPP.

The Gaston data suggest that school personnel were able to effectively diagnose reading achievement as a relative weakness and prioritize this area of instruction. The data clearly indicate a change in direction when compared to prior achievement gains.

Of notable interest at KIPP are the achievement gains attained for a population of students that have traditionally been defined as underperforming. The demographic data suggest that although most of the students in the schools are living in poverty, these same students are significantly increasing levels of achievement.

It is also notable that KIPP schools have decreased the number of students with testing exemptions or missing data, thereby increasing the number of students taking the exams over previous non-KIPP years.

The findings in this report suggest that the three new KIPP schools have been able to successfully repeat the academic success of the two original KIPP schools as measured by academic test scores. Further longitudinal studies are recommended to support the preliminary findings reported in this analysis.

Appendix A
Description of Test Score Data Provided by School and by Grade

	Grade 3 (Pre-KIPP)		Grade 4 (Pre-KIPP)		Grade 5 (KIPP)	
	Stanford 9	State CRT	Stanford 9	State CRT	Stanford 9	State CRT
KIPP DC KEY Academy	None	None	None	None	Intermediate 1 fall (pretest) • Total Reading NCE/SS • Reading Vocabulary NCE/SS • Reading Comprehension NCE/SS • Total Math NCE/SS • Math Problem Solving NCE/SS • Math Problem Solving NCE/SS Intermediate 2 spring (posttest) • Total Reading NCE/SS • Reading Vocabulary NCE/SS • Reading Comprehension NCE/SS • Total Math NCE/SS • Math Problem Solving NCE/SS • Math Problem Solving NCE/SS	None
Gaston College Prep	None	End of Year Exam • Math P/NP/SS/PL • Reading P/NP/SS/PL	None	End of Year Exam • Math P/NP/SS/PL • Reading P/NP/SS/PL	None	End of Year Exam • Math P/NP/SS/PL • Reading P/NP/SS/PL
3D Houston	Primary 3 spring • Total Reading NCE/SS • Reading Vocabulary NCE/SS • Reading Comprehension NCE/SS • Total Math NCE/SS • Math Problem Solving NCE/SS • Math Problem Solving NCE/SS • Total Language NCE/SS	TAAS • TLI Scores • P/NP	Intermediate 1 spring • Total Reading NCE/SS • Reading Vocabulary NCE/SS • Reading Comprehension NCE/SS • Total Math NCE/SS • Math Problem Solving NCE/SS • Math Problem Solving NCE/SS • Total Language NCE/SS • Language Mechanics NCE/SS • Language Expression NCE/SS	TAAS • TLI Scores • P/NP	Intermediate 2 spring • Total Reading NCE/SS • Reading Vocabulary NCE/SS • Reading Comprehension NCE/SS • Total Math NCE/SS • Math Problem Solving NCE/SS • Math Problem Solving NCE/SS • Total Language NCE/SS • Language Mechanics NCE/SS • Language Expression NCE/SS	TAAS • TLI Scores • P/NP

P = Passing, NP = Not Pass, NCE = Normal Curve Equivalent, SS = Scale Score, PL = Performance Level

Education Performance Network KIPP Evaluation

Appendix B

Mean NCE change disaggregated by subgroup, 3D

	Total		LEP		Title 1		SPED		Male		Female		AA		Hispanic		White	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total reading NCE gain score 3 to 4	27	-1.4	17	-0.1	23	-1.0	4	2.6	15	-1.4	12	-1.4	8	-4.0	18	0.4	1	-13.6
Vocabulary NCE gain score 3 to 4	27	-3.7	17	-3.2	23	-2.9	4	2.7	15	-3.1	12	-4.4	8	-4.5	18	-2.8	1	-13.9
Reading comprehension NCE gain score 3 to 4	27	0.3	17	1.5	23	0.3	4	1.5	15	-0.6	12	1.3	8	-2.5	18	2.1	1	-11.0
Total math NCE gain score 3 to 4	27	0.9	17	3.3	23	1.9	4	8.3	15	3.6	12	-2.5	8	0.1	18	1.9	1	-10.9
Problem solving NCE gain score 3 to 4	27	-1.2	17	0.7	23	-0.3	4	10.9	15	1.3	12	-4.3	8	-3.5	18	0.7	1	-16.9
Procedures NCE gain score 3 to 4	27	1.8	17	3.6	23	2.3	4	3.7	15	3.8	12	-0.7	8	4.5	18	0.7	1	0.0
Total language NCE gain score 3 to 4	27	7.4	17	10.6	23	7.7	4	13.6	15	7.0	12	7.9	8	2.7	18	9.7	1	2.8
	Total		LEP		Title 1		SPED		Male		Female		AA		Hispanic		White	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Total reading NCE gain score 4 to 5	49	2.3	34	1.7	45	2.4	4	8.2	30	3.8	19	0.1	11	3.4	36	1.7	2	8.6
Vocabulary NCE gain score 4 to 5	49	5.0	34	4.9	45	5.2	4	0.4	30	5.2	19	4.7	11	2.8	36	5.1	2	15.5
Reading comprehension NCE gain score 4 to 5	49	1.5	34	0.8	45	1.4	4	12.8	30	3.5	19	-1.7	11	3.5	36	0.7	2	3.8
Total math NCE gain score 4 to 5	49	2.6	34	2.0	45	2.2	4	7.5	30	4.1	19	0.3	11	4.3	36	2.0	2	4.6
Problem solving NCE gain score 4 to 5	49	2.1	34	1.1	45	1.1	4	8.9	30	4.5	19	-1.7	11	5.0	36	1.2	2	3.1
Procedures NCE gain score 4 to 5	49	1.4	34	1.5	45	1.9	4	3.8	30	1.4	19	1.5	11	1.2	36	1.2	2	6.1
Total language NCE gain score 4 to 5	49	-0.8	34	-1.7	45	-0.8	4	1.2	30	-0.9	19	-0.8	11	1.3	36	-0.8	2	-13.8
Language mechanics NCE gain score 4 to 5	48	1.2	33	1.1	44	1.3	4	6.4	30	1.2	18	1.2	11	2.7	35	1.6	2	-13.8
Language expression NCE gain score 4 to 5	48	-2.3	33	-3.8	44	-2.4	4	-4.2	30	-2.6	18	-1.8	11	0.9	35	-2.5	2	-16.1

Appendix C

Mean Scale Score by Subgroup

Math Scale Score Descriptive Statistics, Gaston

	Total	Male	Female	White	Minority	Title 1	SPED
Grade 3 (Pre-KIPP)	249.96	250.39	249.42	255.50	249.75	248.82	247.12
Grade 4 (Pre-KIPP)	253.88	252.76	255.27	257.75	253.63	252.46	250.78
Grade 5 (KIPP)	260.97	260.56	261.44	267	260.62	259.77	255.82

Reading Scale Score Descriptive Statistics, Gaston

	Total	Male	Female	White	Minority	Title 1	SPED
Grade 3 (Pre-KIPP)	144.27	142.91	146.04	149	144.02	143.43	140.33
Grade 4 (Pre-KIPP)	146.23	145.11	147.53	150	146	145.16	139
Grade 5 (KIPP)	157.49	156.85	158.24	162.5	157.2	156.62	152.64

Education Performance Network KIPP Evaluation

Appendix D

The passing rates presented in the text of the report are based on a sample that includes only those students actually participating in the test. However, passing rates also were calculated employing *all* students, including those who were exempt from the test and those with missing test-score data. These passing rates are presented for Gaston in the first two tables below and for 3D in the third and fourth tables.

Percent of Students at or Above Proficient End-of-Year Exam (Reading), Gaston

	At or above proficient	Below proficient	Exempt/Missing Data
Grade 3 (Pre-KIPP)	55%	27%	18%
Grade 4 (Pre-KIPP)	53%	41%	6%
Grade 5 (KIPP)	93%	7%	0%

Percent of Students at or Above Proficient End-of-Year Exam (Math), Gaston

	At or above proficient	Below proficient	Exempt/Missing Data
Grade 3 (Pre-KIPP)	48%	27%	25%
Grade 4 (Pre-KIPP)	74%	18%	8%
Grade 5 (KIPP)	90%	10%	0%

Percent of Students Passing, Not Passing, and Exempt TAAS (Reading), 3D

	Passing	Not Passing	Exempt/Missing Data
Grade 3 (Pre-KIPP)	57%	9%	34%
Grade 4 (Pre-KIPP)	66%	11%	23%
Grade 5 (KIPP)	80%	11%	9%

Percent of Students Passing, Not Passing, and Exempt TAAS (Math), 3D

	Passing	Not Passing	Exempt/Missing Data
Grade 3 (Pre-KIPP)	54%	14%	32%
Grade 4 (Pre-KIPP)	71%	7%	22%
Grade 5 (KIPP)	94%	1%	5%

Appendix E
Analysis of Exempt/Missing Student Data

The students with exemptions or missing test data in Grade 4 (Pre-KIPP) at Gaston and 3D were tracked. Their Grade 5 performance was computed. The data reveal that 75% of those Gaston students with no data in Grade 4 passed the Grade 5 End-of-Year exam in reading while 67% with no data in Grade 4 passed the Grade 5 End-Of-Year exam in math.

The Texas data reveal that all students with no Grade 4 data passed the Grade 5 TAAS exams in reading and math.

Gaston Exempt/Missing	
Number E/M (Read)	4
Grade 5 Passing Rate	75%
Number E/M (Math)	6
Grade 5 Passing Rate	67%
3D Exempt/Missing	
Number E/M (Read)	11
Grade 5 Passing Rate	100%
Number E/M (Math)	12
Grade 5 Passing Rate	100%

Glossary of Terms

a priori hypothesis: A term used to describe a hypothesis posed prior to running the statistical analysis.

Completely within-subjects: A term used to indicate that no comparison group was used in a statistical analysis.

State criterion-referenced test (CRT): A test designed to measure the state content standards and report a student's score in relation to the content of the test.

Cross-sectional: Comparing two different groups of students. For example comparing the mean Grade 5 NCE in 2002 with the mean Grade 5 NCE in 2001 measures the same grade, but these are different groups of students at different points in time.

Doubly Multivariate Analysis of Variance (MANOVA): A statistical test with more than one dependent variable (e.g., reading, math, and language scores) measured at more than one point in time (e.g., Grade 3 and 4 and 5).

Eta Squared (η^2): An effect size measure that reports the strength of association. This measure ranges from a low of 0 to a high of 1.

Gain score (change score): Computed by finding the difference between two test scores. For example, a student with an NCE score of 20 in Grade 3 and 30 in Grade 4 would have an NCE gain score of 10.

Interrupted time-series: This is a research design that recognizes a significant change of events, often referred to as an interrupt. For example, the number of drunk driving cases may be measured before and after the implementation of a new state law, where the state law would be considered the interrupt.

Longitudinal: Tracking individual progress over time. For example, computing the gain score for each student in a school over two years is considered a longitudinal approach.

Mean: The arithmetic average of a group of scores.

Normal educational growth: A term used to describe students that were able to maintain or exceed their pretest NCE score. For example, a student with an NCE of 30 in Grade 5 and 28 in Grade 6 did not make normal growth as their posttest NCE was lower than their pretest NCE.

Norm group: A national sample of students participating in the Stanford 9.

Normal Curve Equivalent (NCE) Score: A score reported on standardized achievement tests used to describe a student's score in relation to the distribution. NCEs range from 1 to 99 with a standard deviation of 21.06.

Passing rate: Passing rates described throughout the text are based on those students with test scores. The passing rates in Appendix D are based on all students in the school, whether they actually took the test or not. The latter approach represents the KIPP method of calculation, and provides the percentage of students who were exempt from testing or whose scores were missing.

Performance Levels: These are scores generally reported on criterion-referenced tests representing a student's level of mastery over defined content. These scores are generally broad categories of achievement and should be considered a gross measure of a student's academic performance.

Scale Scores: A score placed on an extended continuum connecting all levels and forms of a testing system. This score allows one to compare student scores on different levels of a test (i.e., Grade 3 and 4 are different levels) and different forms that may be created.

Statistically Significant: Statistical significance implies that the difference between two or more analytical results is truly different. Often scores differ slightly as a result of sampling error, but are considered similar. If a group's mean score is significantly higher than the previous scores, then the difference is not due to sampling error and is likely to represent a real difference. Effect size measures help illustrate how large this difference may be.

Quasi-Longitudinal: Computing the difference between two cohorts over time. For example, computing the difference between the mean Grade 4 NCE in 2001 and the Grade 5 NCE in 2002. This follows the same group of students over time, but is subject to error as some students may have left the school (district) while some new students may have entered the school.