CLARK COUNTY SCHOOL DISTRICT STUDY OF THE EFFECTIVENESS OF SCHOOL IMPROVEMENT PLANS (SESIP) – YEAR TWO

Dr. Kenneth E. Fernandez

I. Introduction

It has been argued that the development of a school improvement plan is an "integral part of every successful ongoing individual school improvement effort."[1] Like many school districts, the Clark County School District (CCSD) is required by state statute (NRS 385.357) or district regulations to develop a school improvement plan (SIP). As mentioned in the CCSD first year report on the effectiveness of the school improvement planning process, CCSD provides an ideal environment to study the effects of planning on school performance. As the fifth largest school district in the nation, planning and evaluation of programs is critically important to increasing achievement and closing the gap. The research division of the CCSD has access to a large amount of individual level data on student achievement, student level socio-economic and demographic information, school level demographic and academic data, and most importantly, data on the quality of each school's SIP. The breadth and quality of the data allow for a rigorous examination of the relationship between a school's SIP and school performance. In the year one report, the results of statistical analyses provided strong evidence that SIP quality was positively and significantly related to school performance. This was true even when controlling for various other factors, or whether one uses various measures of school performance. The year two (2006-2007) report will examine if the positive relationship between the quality of school SIPs and performance is still observed in subsequent years.

While the recent literature suggests that quality SIPs should lead to increases in positive school level outcomes, there may be reasons why one might predict that the relationship between a SIP and school performance would diminish over time. At an individual level, for example - if a person was having health problems due to poor diet and lack of exercise, changes in these activities will likely produce noticeable results. Although continuing this activity will be critical in helping maintain the new "healthy" status quo, it is unlikely that the improvements would increase at the same rate, or that it would not reach some sort of "threshold". At some point, any "improvement" will fail to produce further significant improvements without changes in process or benchmarks measured. Similarly, the same might be said about a new school level program or policy, like a SIP. A SIP may be successful in generating improvements, but it will not produce improvements indefinitely without changes. There are mathematical limitations that would prevent infinite improvement. For example, if a program brought students to the 90th percentile in all test scores there would be little room for improvement thereafter.

Having said this there are reasons one would still observe a correlation between the quality of a SIP and aggregate test scores. Like any good diet or exercise program, it works as long as you stick to it. At an individual level, we might see variation in health improvement depending on the level of commitment to the program. Similarly, over time we will see variation in quality of SIPs at an individual school. Sometimes a school might produce a quality SIP, sometimes it

[1] Doud, James L. 1995. Planning for School Improvement: A Curriculum Model for School Based Evaluation. *Peabody Journal of Education* 70, 3 (Spring): 175-187.

might not. What this produces in the aggregate is an observable amount of variation. If all schools produce similar high quality SIPs then the variation will diminish (and any correlation between quality of SIPs and any other variable will diminish). But, if some schools fluctuate in their quality of SIPs, this observed variation can be matched with variation in school performance to measure the correlation between SIPs and performance.

II. SIP Quality at CCSD

An examination of the SIPs over a three year period shows that overall, quality of SIPs is remaining stable (as measured by the SIP rubric). Table 1 shows that the overall average scores (on 17 factors, each scored between 1 and 3) are very similar for each year, and the amount of variation around the average is also remaining stable. Schools range from poor (scoring the lowest on all 17 dimensions) to excellent (44 out of 51 points).

Table 1: Descriptive Statistics

	SIP Scores (2006)	SIP Scores (2005)	SIP Scores (2004)	DIFFERENCE (2006 - 2004)
N	320	303	277	270
Mean	31.98	29.72	32.11	.34
Median	32.0	30.00	33.00	.00
Mode	32.00	28.00	34.00	-3.00(a)
Std. Deviation	5.16	4.70	5.052	5.47
Variance	26.59	22.12	25.53	29.96
Minimum	18.00	18.00	17.00	-14.00
Maximum	44.00	40.00	44.00	15.00

One of the most important questions is, "Are the same schools consistently producing the same quality SIPs?" That is, are the same schools producing high quality SIPs year after year? Or are the same schools producing lower quality SIPs year after year? The answer is no. One might suspect that certain schools might consistently produce high quality SIPs and consistently score highly on the SIP scoring rubric (see Reeves, 2006). However, this is not the case. First, the correlation between a school's prior year score is modest (Pearson's r of .471 or lower), which means that just because a school received a high quality SIP score the previous year doesn't mean that school will have a high quality SIP the next year. Second, there was no correlation between socioeconomic variables such as "percent of students who are minority" or "per student expenditures" and SIP scores.

This is a positive finding because it means that any school - regardless of percent of student population considered from low income households, distribution of subgroups, or percent of students that are English language learners, among other factors – can produce a high quality SIP. On the other hand, it is discouraging because even if a school has produced an excellent SIP in the past, it does not mean they will continue to do so in the future.

Table 2: Correlations between SIP scores (2004-2006)

		2006 SIP score	2005 SIP score	2004 SIP score	Per Student Expenditures	Percent Minority
2006 SIP score	Pearson Correlation	1	**.471(**)**	**.356(**)**	.046	.057
	Sig. (2-tailed)		.000	.000	.442	.333
	N	320	296	270	286	296
2005 SIP score	Pearson Correlation	**.471(**)**	1	**.440(**)**	-.094	-.022
	Sig. (2-tailed)	.000		.000	.109	.698
	N	296	303	277	292	300
2004 SIP score	Pearson Correlation	**.356(**)**	**.440(**)**	1	**-.130(*)**	-.057
	Sig. (2-tailed)	.000	.000		.031	.345
	N	270	277	277	272	274
Per Pupil Expenditures	Pearson Correlation	.046	-.094	**-.130(*)**	1	**.225(**)**
	Sig. (2-tailed)	.442	.109	.031		.000
	N	286	292	272	295	295
Percent Minority	Pearson Correlation	.057	-.022	-.057	**.225(**)**	1
	Sig. (2-tailed)	.333	.698	.345	.000	
	N	296	300	274	295	307

* Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

III. The Relationship between SIP Quality and Average School Achievement, 2006-2007

In this analysis, two measures were used to represent student achievement by school. The Iowa Test of Basic Skills (ITBS) and the Iowa Test of Educational Development (ITED) were the norm referenced tests used for the first part of the analyses. A second measure, the Nevada Criterion Referenced Test (CRT), was also used.

A. Norm Referenced Tests – ITBS and ITED

To examine the influence SIPs have on school performance changes, a measure of school performance is needed. Results from the ITBS (grades 4 - 8) and the ITED (grades 9-10) norm referenced tests (NRTs) were used for each student who took the test in the 2006-2007 school year. One advantage of using the NRTs is that the tests are vertically aligned (comparable). This means that a student growth score can be obtained by simply subtracting the prior year's test results with the current test. (NRT score in 2006-2007 minus the NRT score in 2005-2006). This was done for both math and reading scores for students who were in the same school in 4th grade in 2007 as they were in 3rd grade in 2006, and students who were in the same school in 7th grade as they were in 6th grade, and students who were in the same school in 10th grade as they were in 9th grade. The individual student growth scores for math and reading were then aggregated by school to create an average school performance change score for both math and reading.

Table 3: CCSD Average Change Score 2006 - 2007

	N	Minimum	Maximum	**Average Change 2005-2006**	**Average Change 2006-2007**	Std. Deviation
Average Change in Reading Score	287	-14.25	33.00	**13.4**	**14.32**	5.55
Average Change in Math Score	286	-10.33	26.41	**13.5**	**14.07**	6.71

The range of improvement (change) in average school NRT standard scores (or lack there of) was similar to last year's range. This year's average school NRT improvement was a little higher than last year's (which was an average 13.5 point improvement in mathematics scores and 13.4 points in reading).

School Performance Analyses

The two measures of NRT average gain score were used as the dependent variables in the Ordinary Least Square (OLS) regression. The main independent (explanatory) variable of interest in the regression model was a measure of the quality of a school's SIP. Although all schools were guided by a general planning format/framework and were following district and state policies, substantial variation was found in the quality of SIPs among CCSD schools. To measure the quality of SIPs, CCSD, along with the Center for Performance Assessment, created a scoring rubric that grades the quality of a school improvement plan. This rubric was used by the Center to determine a "quality" score for each of the 330 SIPs based on 17 different indicators (see appendix for a description of each). Each dimension was given a score that ranged from 1 (lowest rating) to 3 (highest rating).

OLS regression statistically controls for correlation among independent variables in order to get an unbiased estimate of the effect of each independent variable on the dependent variable. By controlling for other potential explanations (other independent variables), we can be more sure that one variable is associated with a change in another variable by isolating the contribution that SIP quality has on school performance. Several control variables are used to make sure that the relationship between SIP quality and school performance is not spurious (actually caused by some other factor). Percent minority, percent eligible for free and reduced lunch (FRL), percent Limited English Proficiency (LEP), percent of student with an Individual Education Plan (IEP), grade level (high school measured as a dichotomous variable), transiency rate, and per capita school spending were used as the control variables in the OLS regression models. OLS regression is the ideal technique to analyze the average NRT math and reading improvement scores for each school. OLS works best when the dependent variable is normally distributed (bell shaped). Both improvement measures are normally distributed (See figure 1 and 2).

Figure 1: Histogram of Reading Improvement Score

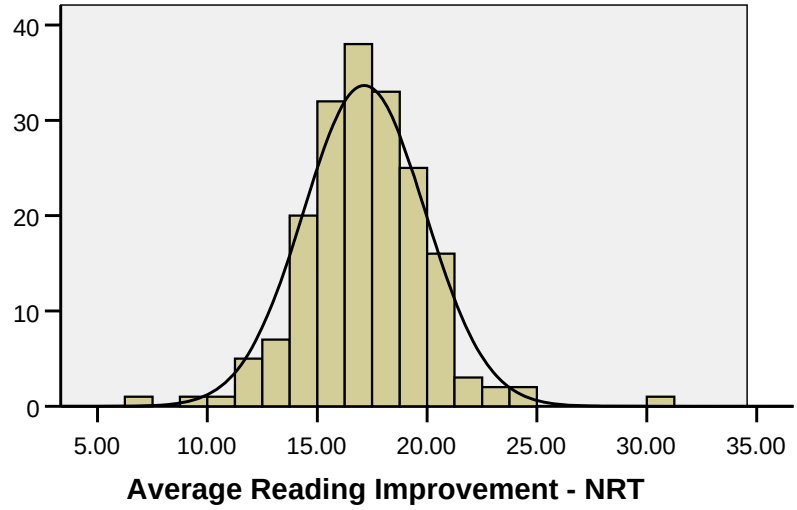
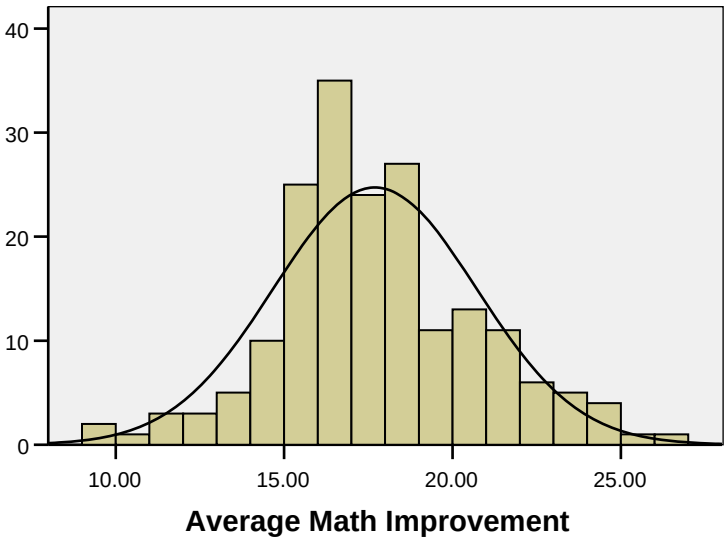


Figure 2: Histogram of Math Improvement Score



This suggests that this type of measure is capturing the variation in school performance and that OLS regression is the appropriate technique to study what factors influence school performance (especially the quality of a school's SIP).

Tables 4 and 5 show the results of an OLS regression analysis of average school improvement on the NRT for math and reading. The dataset contains scores for 287 schools (187 Elementary schools, 58 Junior High schools, and 42 High schools).

Table 4: OLS Regression – All Schools - Improvement on NRT Reading

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t-value	Significance
	B	Std. Error	Beta		
(Constant)	17.035	1.719		9.910	.000
SIP Score 2006	**.110**	**.046**	**.096**	**2.403**	**.017**
Percent LEP	8.677	1.818	.210	4.774	.000
Percent Minority	-3.211	1.017	-.159	-3.159	.002
Percent IEP	2.699	1.490	.072	1.812	.071
Percent FRL	-8.358	1.031	-.339	-8.108	.000
Transiency Rate	-.092	.015	-.268	-6.165	.000
Per Pupil Expenditure	.123	.104	.047	1.177	.240
High School	-8.360	.609	-.553	-13.736	.000

R	R Square	Adjusted R Square	Standard Error of the Estimate	# of Schools
.811	.657	.646	2.94968	267

Table 5: OLS Regression – All Schools - Improvement on NRT Mathematics

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	13.341	1.996		6.685	.000
SIP Score 2006	.129	.053	.086	2.425	.016
Percent LEP	8.799	2.110	.162	4.170	.000
Percent Minority	-2.477	1.180	-.093	-2.099	.037
Percent IEP	4.877	1.729	.099	2.820	.005
Percent FRL	-9.959	1.197	-.307	-8.321	.000
Transiency Rate	-.035	.017	-.077	-2.005	.046
Per Pupil Expenditure	.318	.121	.093	2.627	.009
High School	-13.551	.707	-.681	-19.180	.000

R	R Square	Adjusted R Square	Standard Error of the Estimate	# of Schools
.856	.733	.724	3.42458	267

The results of the statistical analysis were consistent with the prior year's SIP study and confirmed that the quality of a school's improvement plan is strongly correlated with school academic performance as measured by average student improvement on the norm referenced tests. This positive association remains statistically significant even when controlling for a variety of school factors, including per-pupil spending, percent of students who are a minority, percent limited English proficiency (LEP), and percent of students eligible for free or reduced lunch (FRL).

B. Criterion Referenced Tests (CRT)

The above regression models could not be applied in the same way for the Criterion Referenced Tests. Because the CRT is not designed to be vertically aligned (comparable), calculating a student growth score would not produce an accurate measure of growth. The aggregate average scores for each school can still be used as a measure of school performance, but cannot be pooled together (i.e., combine elementary, junior high, and high schools in the same dataset). Each type of school needs to be analyzed separately which limits the sample size and therefore the statistical power of any model.[2] Even with those problems that dampen the ability of a statistical tool to uncover a relationship between two variables, **there can be seen a statistically significant correlation between the quality of a school's SIP and school performance.** For example, a correlation analysis (see table 6) of 8th grade average CRT scores and SIPs found a positive and statistically significant relationship.

Table 6: Correlations between CRT Average Scores and SIP Scores

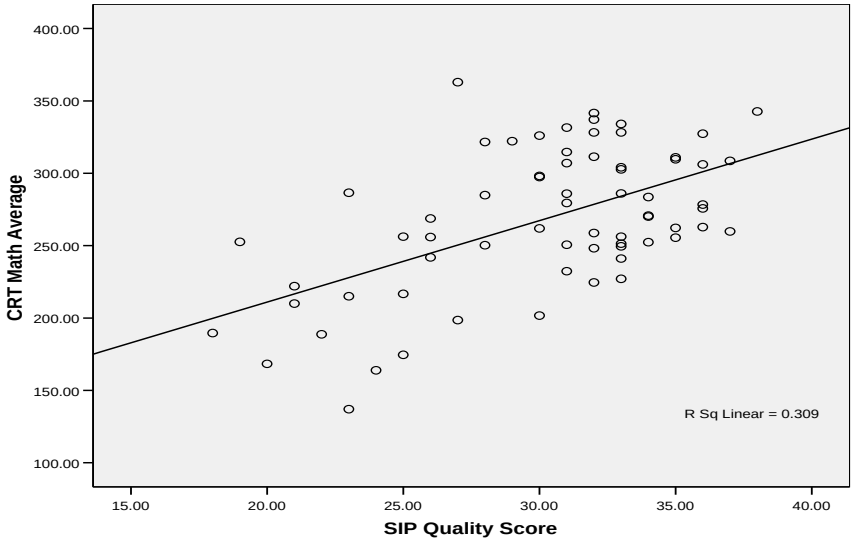
		CRT Mathematics	CRT Reading	SIP Score 2006
CRT Mathematics	Pearson Correlation	1	**.912(**)**	**.555(**)**
	Sig. (2-tailed)		.000	.000
	N	73	73	67
CRT Reading	Pearson Correlation	**.912(**)**	1	**.530(**)**
	Sig. (2-tailed)	.000		.000
	N	73	73	67
SIP Score 2006	Pearson Correlation	**.555(**)**	**.530(**)**	1
	Sig. (2-tailed)	.000	.000	
	N	67	67	320

** Correlation is significant at the 0.01 level (2-tailed).

This positive and significant relationship can perhaps be better seen in a plot of the two variables (CRT score and SIP score). Figure three visually describes this relationship using 8th grade CRT math averages for each middle school.

[2] Statistical power is the probability of obtaining statistical significant results when the hypothesis being tested is true.

Figure 3: Scatter plot of CRT Math Scores and SIP Scores



The scatter plot clearly shows that as the quality of a SIP increases so does school performance as measured by CRT mathematics scores. In addition, bivariate OLS regressions were performed on every grade where CRT data was available (3rd through 8th). In this analysis current CRT averages (the dependent variable) were compared to last year's SIP quality (2005) scores (the independent variable). **In every grade level, for both reading and math, there was a statistically significant relationship between last year's SIP quality score and this year's school performance (as measured by CRT averages for math and reading). This suggests that in addition to the strong relationship between SIP quality and school improvement in last year's study, there may also be a residual effect where prior SIP quality is moderately influencing future school performance.**

IV. Conclusions/Recommendations

The relationship between good planning and performance is something that interests many organizations are interested in. Whether it is a large private corporation or a small public school, good planning seems to make intuitive sense. But there are scholars and practitioners who have stated that planning exercises are a waste of time and resources and perhaps even detrimental to performance.[3] The findings in this study provide strong evidence of a consistent and robust relationship between school improvement plans and school performance. The results are quite consistent with last year's examination which also found that SIP quality is positively and significantly related to school performance, even when controlling for various other factors, or whether one uses various measures of school performance. The two studies combined suggest several things. First, schools that are producing high quality SIPs are not just the upper income level area, homogeneous schools, but rather any school has the capacity to write a good school improvement plan. Second, just because a school produced a high quality SIP previously doesn't mean that same school will produce a high quality SIP the following year. However, when a

[3] For example see Henry Mintzberg's 1994 book The Rise and Fall of Strategic Planning. New York: Free Press.

school does produce a high quality school improvement plan it is usually followed by strong academic performance. It also suggests that simply mandating schools to produce SIPs is not enough, but that: schools need guidance in the process of developing and maintaining high quality SIPs; SIPs need to be evaluated for quality yearly; new principals or principals new to schools need training and guidance in the SIP process; and schools need to receive feedback on the quality of their SIP and know why they were scored high or low.

APPENDIX

The 17 dimensions used to evaluate the SIPs include:[4]

1. Comprehensive – all goals are linked to identified concerns and causes
2. Specific Goals – targeted students and subgroups, targeted standards
3. Measurable Goals – quantifiable goals with a baseline measurement and a target
4. Achievable Goals – goals are sufficiently challenging to close learning gaps in 3-5 years for targeted subgroups
5. Relevant Goals – all goals align with urgent student needs identified in comprehensive needs assessment
6. Timely goals – all goals identify a specific window of time when the assessment will be administered
7. Inquiry Process – all causes explicitly aligned with solutions
8. Research-based strategies – all strategies/solutions address standards-based research strategies
9. Master plan design – action steps consistently describe how solutions will be implemented
10. Professional development gaps – all program implementations are supported by specific action steps for professional development
11. Professional development focus – evidence that the professional development will be sustained and incorporated into routine operations within the school
12. Parental involvement – evidence of frequent parent communication regarding standards, best practices, and grading
13. Monitoring plan – monitoring steps explicitly describe what people are doing to assess progress toward goal attainment
14. Monitoring frequency - 3 = weekly; 2 = monthly or more; 1 = less than monthly
15. Evaluation process – measures selected allow planned outcomes to be compared with achieved outcomes; evaluation plan explicitly describes how action will be taken as a result of the evaluation plan
16. Use of time and opportunities – described strategies address time and opportunity to meet needs of targeted subgroups; document how existing time and opportunity will be improved
17. Other required elements of a SIP

[4] For a more detailed discussion of the SIP scoring rubric see the report "Achievement, Equity, and Leadership: the Clark County School District Commitment to Success" by the Center for Performance Assessment (2006).