Effects of a Data-Driven District-Level Reform Model

Robert E. Slavin
Johns Hopkins University

GwenCarol Holmes
Nancy A. Madden
Anne Chamberlain
Success for All Foundation

Alan Cheung
Johns Hopkins University

February, 2010

This research was supported by a grant from the Institute for Education Science, U.S. Department of Education (No. R-305A040082). However, any opinions expressed are those of the authors and do not represent the positions or policies of the U.S. Department of Education.

Effects of a Data-Driven District-Level Reform Model

Abstract

Despite a quarter-century of reform, US schools serving students in poverty continue to lag far behind other schools. There are proven programs, but these are not widely used. This large-scale experiment evaluated a district-level reform model created by the Center for Data-Driven Reform in Education (CDDRE). The CDDRE model provided consultation with district leaders on strategic use of data, identification of root causes, and selection of proven programs capable of solving identified problems. A matched time series design compared trends on state reading and math tests after introduction of CDDRE services. After three years, significant positive effects were found on elementary and middle school reading but not mathematics. The findings support the idea that districts can improve outcomes using a data-driven reform process, but they need to implement changes in light of the data to see benefits for student achievement.

For at least a quarter century, schools in the US have been in a constant state of reform. Commission reports, white papers, politicians, and the press periodically warn of dire consequences if America's schools are not substantially improved. In fact, on the National Assessment of Educational Progress (2007) and on some international measures such as TIMSS (2007), PISA (2006), and PIRLS (2006), US schools have shown some gains in recent years, but the pace of change is slow. In particular, although the academic performance of middle class students is comparable to that of similar students in other countries, the most important problem in the US is the continuing low achievement of disadvantaged and minority students. For example, on the NAEP (2007), 43% of White students scored proficient or better, while only 14% of African American, 17% of Hispanic, and 18% of American Indian students scored at this level. Among students who do not receive free lunches, 44% scored at proficient or better, while among those who receive free lunches, only 17% scored at proficient or better. Results in math and at different grade levels showed similar gaps.

The continuing low performance of disadvantaged and minority students must be considered in light of substantial evidence showing positive effects of a wide range of educational innovations. Many interventions have been evaluated in rigorous experiments and found to improve student achievement, especially in reading and math, in comparison to traditional methods. Yet programs with strong evidence of effectiveness are rarely widely used, and those that are widely used rarely have much, if any, evidence of effectiveness. For example, there were five commercial reading texts that were emphasized in the Reading First program and were among the most widely used in the US during the period from 2000 to the present. The What Works Clearinghouse (2009a), in its beginning reading review, found supportive evidence for none of them. The same lack of evidence for these programs was reported in a review by Slavin, Lake, Chambers, Cheung, & Davis (2009). Reading programs that did have evidence of effectiveness from rigorous evaluations, such as tutoring, cooperative learning, and comprehensive school reform, are not used widely enough to have any meaningful impact on the national achievement gap. The same disconnect exists in math, where widely used textbook and CAI programs have little evidence of effectiveness (What Works Clearinghouse, 2009b, c; Slavin & Lake, 2008; Slavin, Lake, & Groff, 2009), while programs that do have extensive evidence of effectiveness are not widely used.

The limited application of proven programs is perhaps surprising in light of the extraordinary pressure schools have been under in recent years to improve student achievement. Under No Child Left Behind, schools are subject to increasing sanctions leading up to closure or reconstitution if they do not meet standards on state accountability measures for a period of years. Because of the universal availability of data on student performance and the pressure to increase scores, it might be assumed that schools and districts would be intent on finding and adopting programs with strong evidence of effectiveness on the types of measures for which they are held accountable. Yet this is rarely the case.

Data-Driven Reform

The push to improve test scores has led to substantial interest in the use of data within schools and districts to drive decisions and motivate change. The focus of data-driven reform approaches is on obtaining timely, useful information, trying to understand the "root causes" behind the numbers, and designing interventions targeted to the specific areas most likely to be inhibiting success. The idea is both to focus resources and efforts most efficiently where they will make the biggest difference and to break the daunting task of turning around entire schools and districts into smaller achievable tasks that can be accomplished in a reasonable time period, building a sense among front-line educators that they are capable of making a difference on enduring problems.

Data-driven reform involves collection, interpretation, and dissemination of data intended to inform and guide district and school reform efforts. Bernhardt (2003) identified four categories of data districts may analyze: student learning, demographics, school process, and teacher perceptions. These enable school leaders to identify specific problems faced by students and teachers, to break down the data to identify individual schools and demographic groups in need of particular help, and to suggest reasons for achievement gaps (Kennedy, 2003; Schmoker, 2003). Data-based decision making usually involves extensive professional development for school leaders to help them use data to set goals, prioritize resources, and make intervention plans (Conrad & Eller, 2003).

There is surprisingly little evidence on the effectiveness of data-driven reform strategies. That which does exist consists primarily of case studies of schools or districts that have made significant progress on state assessments. For example, the Council of the Great City Schools (2002) identified big-city districts that consistently "beat the odds" in raising student achievement, concluding that these districts were characterized by coherence, planfulness, and extensive use of data to inform district and school decisions. Case studies of other "positive outlier" districts and states have reached similar conclusions (Snipes, Doolittle, & Herlihy, 2002; Grissmer & Flanagan, 2001; Streifer, 2002; Symonds, 2003; CCSSO, 2002). However, such case studies only provide after-the-fact explanations of good results. We do not know, for example, whether schools and districts that did not make impressive gains may also have been trying to use the same data-driven strategies (see Herman et al., 2008).

Frequently, districts embarking on data-driven reform adopt benchmark assessments given three to five times a year to determine whether students are on track toward improvement on their state assessments. The idea is to find out early where problems may exist so that changes can be made before it is too late. Sensible as this appears, there are no experimental evaluations of the use of benchmark assessments. There is evidence that more frequent assessment is more effective than annual assessment (e.g., Dempster, 1991; Schmoker, 1999; Bangert-Drowns et al., 1991), but this is not the same as direct evaluation of any particular application of benchmark assessments.

Center for Data-Driven Reform in Education

In 2004, the US Department of Education funded a research center at Johns Hopkins University to create and evaluate a replicable approach to whole-district change based on the concepts of data-driven reform. The Center for Data-Driven Reform in Education (CDDRE) was intended to try to solve the problem of scale in educational reform by working with entire school districts. The idea was to help district and school leaders understand and supplement their data, identify root causes underlying important problems, and then select and effectively implement programs directed toward solving those problems. The program was initially a partnership between Johns Hopkins and several non-profit organizations that provide training and materials to support whole-school turnaround and have good evidence of effectiveness: Success for All (Slavin, Madden, Chambers, & Haxby, 2009), Direct Instruction (Adams & Engelmann, 1996), America's Choice (Supovitz, Poglinko, & Snyder, 2001), Modern Red Schoolhouse (2002), and Co-nect (Russell & Robinson, 2000). All of these were found to have good evidence of effectiveness by the Comprehensive School Reform Quality Center (CSRQ, 2006).

Best-Evidence Encyclopedia

In addition to whole-school reform models, CDDRE offered information to schools and districts on reading and math programs with strong evidence of effectiveness. Initially, it was expected that reviews of the evidence on such programs would soon be forthcoming from the What Works Clearinghouse, but the WWC reviews did not appear in time, so CDDRE created its own set of reviews, called the Best-Evidence Encyclopedia (BEE; see www.bestevidence.org). They eventually covered elementary math (Slavin & Lake, 2008), secondary math (Slavin, Lake, & Groff, 2009), elementary reading (Slavin, Lake, Chambers, Cheung, & Davis, 2009), and secondary reading (Slavin, Cheung, Groff, & Lake, 2008).

The CDDRE Intervention

The services provided by CDDRE were designed to help district leaders understand and manage their own data, identify key areas of weakness and root causes for these deficits, recognize strengths and resources for reform, and then select and implement programs with strong evidence of effectiveness targeted to their identified areas of need. CDDRE consultants, all of whom had experience as superintendents, principals, or other leadership roles in education, provided approximately 30 days of on-site consultation to each district over a two-year period, depending on district size.

Data Review. CDDRE consultants cooperatively planned a series of meetings with district leaders and school teams (principal and key staff) to engage in a process of exploring all sources of data already collected by the district, including standardized test scores, attendance, disciplinary referrals, retentions, special education placements, and dropouts. CDDRE consultants and district leaders discussed the district's experiences with reform programs already in place, resources, state and federal mandates and constraints, and other factors relevant to the

district's readiness for reform. Surveys of teachers collected information on their perceptions of school strengths and needs.

Benchmark Assessments. Under a subcontract to CDDRE, the Success for All Foundation created a set of state-specific benchmark assessments that assessed reading and mathematics achievement in grades 3-8 (in Pennsylvania, grades 3-11). These quarterly benchmark assessments, called 4Sight, were created from the same assessment blueprints as those used to construct the state assessments, and were written to mirror the state assessment's content, coverage, difficulty, item types, proportions of open-ended items, and use of illustrations and other supports. The 4Sight benchmarks correlated with scores on the state test in the range of +0.80 to +0.85. 4Sight benchmarks were used 4-5 times per year to predict what students, student subgroups, classes, and schools would have scored on the state assessments. Special software enabled school leaders and teachers to examine the data by state standard, grade, class, student subgroup, and so on. The benchmark assessments provided district and school leaders with detailed, timely, actionable information on student achievement, giving them an opportunity to take action in time to affect yearly outcomes.

School Walk-Throughs. CDDRE consultants accompanied district leaders on visits to a cross-section of the district's elementary, middle, and high schools. These structured walk-throughs provided insight for both the CDDRE consultants and the district administrators into the quality of instruction, classroom management, motivation, and organization of each school. They examined the implementation of various programs the schools were using, and focused on student engagement. In addition to informing CDDRE consultants, these walk-throughs were useful in helping district leaders understand the real state of education in their own schools, to find out which of the many programs provided to their schools were actually in use, and to create a sense of urgency to take action.

Data-Based Solutions. Although many of school leaders believed that the knowledge provided by benchmark assessments, data reviews, and walk-throughs were sufficient to cause reform to take place, the CDDRE model emphasized the idea that systematic reforms based on the data are essential if genuine progress is to be made. CDDRE consultants helped district and school leaders review potential solutions to the problems they identified. They emphasized programs and practices with strong evidence of effectiveness, those identified by the BEE or the What Works Clearinghouse. CDDRE consultants helped district and school leaders learn about research-proven solutions, and then advised them through a process of adopting and implementing them: obtaining teacher buy-in, ensuring high-quality professional development and follow-up, and doing formative assessments of program outcomes.

Focus of the Evaluation

The evaluation of the CDDRE process was intended to determine the value added to student achievement by the intervention throughout the districts involved. The intervention was delivered over a period of years, and had distinct components at different points in time that were expected to affect outcomes differentially. In the first year, all participating districts received

extensive consulting on data-driven reform and almost all implemented benchmark assessments (unless they were already in use). Early-years outcomes therefore were exclusively evaluations of the data interpretation aspects of CDDRE. In later years, as schools began to select and then implement proven programs, outcomes begin to reflect the effects of these programs. It was not the intention of the evaluation to examine impacts of particular programs, but rather to focus on the impact across the districts of the process that led to the selection and implementation of proven programs attuned to their needs. Since schools that implemented programs did so at different times in different subjects, the effects of the process would be expected to appear gradually over time.

The original design of the CDDRE intervention involved random assignment of pairs of similar districts within states to experimental or control conditions. A total of 39 districts in five states [1] (PA, AZ, MS, IN, OH) were recruited and randomly assigned in this way over a 3-year period. In order to facilitate recruitment, a delayed treatment control group design was used, in which districts assigned to the control groups received the full treatment a year later. This design would have been appropriate if the interventions had been adopted quickly and completely. However, this was not the case. Implementation was slow and uneven across the districts. Because the intervention was implemented in the "control groups" in the second year, the "control groups" were often as far along as their matched "experimental" groups by the end of the experimental groups' second year. As a result, the randomized experimental-control design was abandoned. Instead, all of the selected districts were considered experimental groups, starting on whatever date they began to receive CDDRE services. Control groups composed of schools that had never been involved with CDDRE were chosen from each state to match CDDRE districts in terms of prior state test scores, demographics, and other factors, as presented in the following section.

This allowed us to follow districts over time as they incorporated the CDDRE elements and compare outcomes to those of schools in districts as similar as possible the experimental districts except for the potentially important fact that the experimental districts volunteered to participate in the experiment and the control districts did not.

The research question was as follows:

In comparison to schools that never received CDDRE services, what were the effects of CDDRE participation on state tests of reading and math at the elementary and middle school levels?

Methods

Sample Selection

CDDRE districts were recruited by forming partnerships with state departments of education in five states. The state departments then nominated districts with many low-achieving schools. The leadership of the nominated districts was approached by CDDRE staff and offered

[1] Two additional states participated, but have not provided complete data

the opportunity to participate in the project, understanding that they would be randomly assigned to receive CDDRE services beginning either the following fall or a year later. The districts were recruited in three cohorts, beginning in fall of 2005, 2006, and 2007. Districts were grouped according to the year they began to receive CDDRE services. For example, a district randomly assigned to the delayed treatment group in 2005 and one randomly assigned to the immediate treatment group in 2006 would be in the same 2006 cohort. Within each district, district leaders could designate all schools or a subset of low-achieving schools to receive CDDRE services. Elementary, middle, and high schools participated. Most of the districts (28) were in Pennsylvania, and there were 4 in Arizona, 4 in Mississippi, 2 in Indiana, and 1 in Ohio, for a total of 39 districts. All were high-poverty Title I districts and schools, but they ranged from small rural districts to large urban ones. There were a total of 296 schools that received CDDRE services across all cohorts and states.

The control group was identified from state records in districts not receiving CDDRE services. Matches were made based on prior state test scores in reading and math, percent free/reduced lunch, and percent minority. Initially, the plan was to identify two matches for each CDDRE school to increase statistical power. However, the participating schools were so low in prior achievement and high in poverty levels that in some states there were few matches available. At the fifth grade level, a total of 183 CDDRE and 197 matched control schools were identified. Among schools with an eighth grade, there were 92 CDDRE and 104 control schools. Characteristics of the CDDRE and control schools are summarized in Table 1.

==========

TABLE 1 HERE

==========

As is apparent from Table 1, the CDDRE and control schools served mostly high-poverty, minority students. Sixty-four percent of fifth graders and 67% of eighth graders qualified for free- or reduced-price lunches. Sixty-five percent of fifth graders and 66% of eighth graders were members of minority groups, mostly African Americans. The CDDRE and control groups were significantly different only on eighth grade percent minority ($p<.05$), where the control group had a somewhat higher percentage of minority students. Pretest scores in reading and math were very similar in CDDRE and control schools.

Measures

The measures for this study were the reading and math assessment scores for each state at the 5th and 8th grade levels. These were the Pennsylvania System of School Assessment (PSSA), the Arizona Instrument to Measure Standards (AIMS), the Mississippi Curriculum Test 2 (MCT-2), the Indiana Statewide Testing for Educational Progress-Plus (ISTEP+), and the Ohio Achievement Test (OAT). Standard school-level scores on these measures were taken from each state's website, for the year prior to CDDRE implementation through 2008.

Prior to analysis, all scores were transformed to z-scores within states and grade levels, to permit pooling across states and years. Note that this removes year-to-year variations likely to result from variations in test versions within states. These would affect control and experimental schools equally. A z-score of zero indicates that a school is scoring at the average for its set of matched experimental and control schools in a given year.

Analyses

A simple t-test was used to determine whether there was any significant difference in the achievement levels of treatment and control groups at pretest. There were no significant pretest differences. Following this, analyses of covariance (ANCOVAs) were used to compare the treatment and control groups each year after CDDRE implementation (up to 3 years), using the pre-implementation year's achievement as a covariate. Because schools joined the CDDRE project in successive waves or cohorts, the number of schools available for comparison at each post-test year diminishes over time. That is, while all schools included in the analysis have at least 1 year of posttest data, a smaller number have accumulated 3 posttest years. Effect sizes were computed as the experimental-control difference in adjusted posttest scores divided by the unadjusted school-level standard deviation.

Results

==========
FIGURES 1-4 HERE
==========

The findings are summarized in Figures 1-4. Figure 1 shows outcomes for fifth grade reading. At the end of the first implementation year, CDDRE schools scored significantly lower than control schools (ES= -0.24, $p<.01$). CDDRE schools were nonsignificantly lower than controls in Year 2 (ES= -0.11, n.s.), but in Year 3, CDDRE schools scored significantly higher than controls (ES= +0.40, $p<.01$).

Figure 2 shows reading scores for eighth grade reading. At the end of the first implementation year, CDDRE schools scored slightly worse than controls (ES= -0.04, n.s.), but were significantly higher after the second year (ES= +0.28, $p<.03$) and were marginally higher in the third year (ES= +0.40, $p<.07$).

Fifth grade math scores, depicted in Figure 3, do not show any differences between CDDRE and control schools. In eighth grade, there were no differences in the first year (ES=+0.01, n.s.), but CDDRE schools scored significantly higher in the controls in the second year (ES=+0.27, $p<.05$). In the third year, there were once again no differences (ES=+0.07, n.s.).

Discussion

The findings of the evaluation of CDDRE were mixed, but there were some important trends worthy of note. First, there were no first-year effects in either subject or grade level. Clearly, the provision of workshops and implementation of benchmarks is not sufficient to bring about changes in student performance. This finding is in accord with the program's theory of action; first year interventions are analogous to taking a patient's temperature, not providing a treatment.

By the second year, school and district leaders were, in many locations, beginning to take action based on the data. Perhaps as a result, test scores improved on both measures at both grade levels, although there were significant differences (compared to controls) only on 5th and 8th grade reading. In the third implementation year, many more schools adopted either proven programs or other solutions, particularly in reading, and there were significant and substantial effects in reading at both the fifth and the eighth grade levels. Effects were not seen, however, on math scores. This difference between reading and math outcomes corresponds with the pattern of program adoption by schools that implemented proven models. Across all schools, 65% of programs chosen were in reading, 25% in math, and 10% in "other".

What the findings imply is that helping school district leaders understand student data is not enough to produce gains in achievement. Schools must actually take action to change teaching and learning.

Where CDDRE appeared to make its largest differences, in reading at both grade levels, the magnitude of the effects was surprisingly large, averaging an effect size of +0.40 at both levels. These effect sizes cannot be compared to individual-level effect sizes, because standard deviations (the denominator of the effect size formula) are 2-3 times larger among students than among schools. However, the ability to make this much difference on such a large scale is important. If outcomes of similar magnitude are seen in replications, these findings may point to a relatively inexpensive, readily scalable strategy for making a different in the performance of high-poverty schools.

On the other hand, the changes brought about by participating in the CDDRE process took a long time to appear, and many schools and districts never implemented any systematic changes in teaching in response to the data. According to the experience of this study, districts or schools ready to make reforms in teaching should go straight to them, perhaps choosing among proven models focused on the problems of greatest concern, rather than engaging in a long process of identifying problems and implementing benchmark assessments. Reviews of research on structured comprehensive reform models (see CSRQ, 2006a,b) typically find positive first-year effects on achievement measures among schools that volunteered to use the programs. The CDDRE process may be beneficial for districts not ready to engage in school-level reforms, but nothing in the CDDRE experience or the data supports the idea that focusing school leaders on data will in itself improve outcomes.

References

Adams, G.L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond. DISTAR* Seattle, WA: Educational Achievement Systems.

Bangert-Drowns, R.L., Kulik, C.C., Kulik, J.A, & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61* (2), 213-238.

Bernhardt, V.L. (2003). No schools left behind. *Educational Leadership, 60* (5), 26-30.

Comer, J. P., Haynes, N. M., Joyner, E. T., & Ben-Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education.* New York: Teachers College Press.

Comprehensive School Reform Quality Center (2006a). *CSRQ Center report on elementary school comprehensive school reform models*. Washington, DC: American Institutes for Research.

Comprehensive School Reform Quality Center (2006b). *CSRQ Center report on secondary school comprehensive school reform models*. Washington, DC: American Institutes for Research.

Conrad, W.H., & Eller, B. (2003, April). *District data-informed decision making*. Paper presented at the annual meetings of the American Educational Research Association, Chicago.

Council of the Great City Schools (2002). *Beating the odds II.* Washington, DC: Authors.

Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership, 72*(8), 71–76.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning around chronically low-performing schools: A practice guide* (NCEE #2008-4020). Washington DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/practiceguides.

International Association for the Evaluation of Educational Achievement (IEA) (2003). *Trends in international mathematics and science study (TIMSS), 1995, 1999, and 2003.* Boston, MA: Author.

Kennedy, E. (2003). *Raising test scores for all students: An administrator's guide to improving standardized test performance*. Thousand Oaks, CA: Corwin Press.

Modern Red Schoolhouse (2002). *School improvement: Research results for Modern Red Schoolhouse*. Nashville, TN: Author.

National Assessment of Educational Progress (2007). *The nation's report card.* Washington, DC: National Center for Education Statistics.

Russell, M., & Robinson, R. (2000). *Co-nect retrospective outcomes study*. Boston, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

Schmoker, M. (1999). *Results: The key to continuous school improvement (2nd ed.).* Alexandria, VA: ASCD.

Schmoker, M. (2003). First things first: Demystifying data analysis. *Educational Leadership, 60 (5),* 22-25.

Slavin, R.E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best evidence synthesis. *Reading Research Quarterly*, 43(3), 290-322.

Slavin, R.E., & Lake, C. (2008). Effective programs in elementary math: A best evidence synthesis. *Review of Educational Research*, 78(3), 427-515.

Slavin, R.E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79 (2), 839-911.

Slavin, R.E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79* (4), 1391-1465.

Slavin, R.E., Madden, N.A., Chambers, B. & Haxby, B. (2009). *Two million children: Success for All*. Thousand Oaks, CA: Corwin.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: Council of the Great City Schools.

Streifer, P. (2002). *Using data to make better educational decisions.* Lanham, MD: Scarecrow Press.

Supovitz, J.A., Poglinco, S.M., & Snyder, B.A. (2001). *Moving mountains: successes and challenges of the America's Choice comprehensive school reform design*. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

What Works Clearinghouse (2009a). Beginning Reading Topic Report. At http://ies.ed.gov/ncee/wwc/reports/beginning_reading/topic/

What Works Clearinghouse (2009b). Elementary Mathematics Topic Report. At http://ies.ed.gov/ncee/wwc/reports/elementary_math/topic/ 18

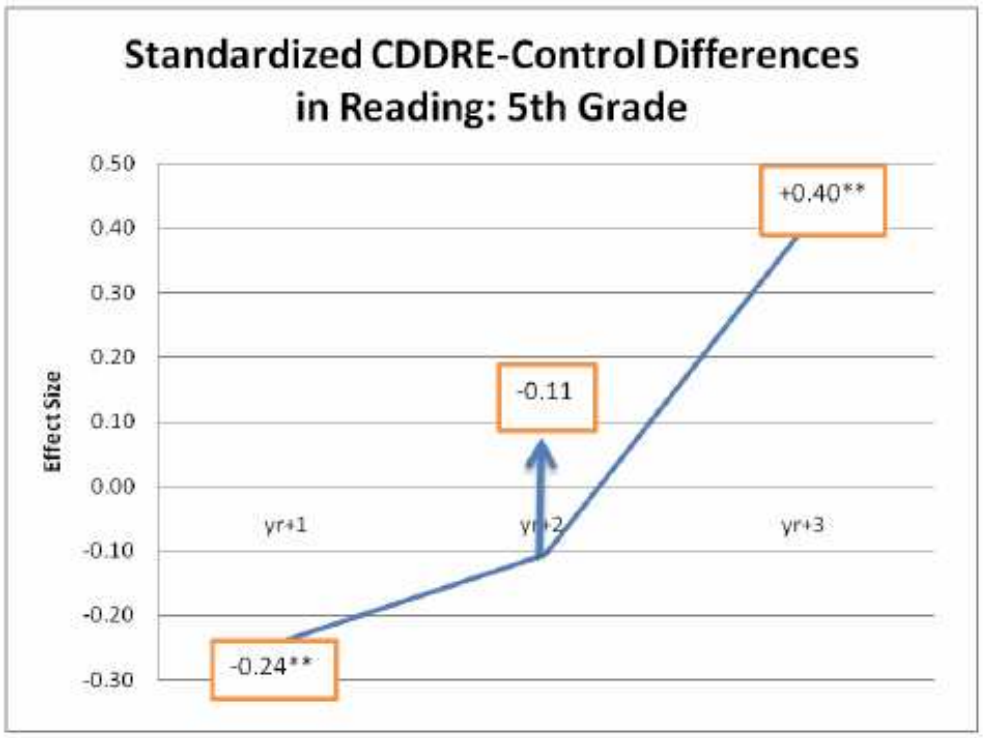
What Works Clearinghouse (2009c). Middle School Mathematics Topic Report. At http://ies.ed.gov/ncee/wwc/reports/middle_math/topic/

Table 1

Demographic and Pretest Characteristics Of Matched CDDRE and Control Schools

		5th Grade	8th Grade
% Free Lunch			
CDDRE	x	62.79	68.59
	(SD)	(34.21)	(31.18)
	N	183	92
Control	x	64.32	66.83
	(SD)	(33.74)	(42.49)
	N	197	104
		5th Grade	8th Grade
% Minority			
CDDRE	x	62.43	67.28
	(SD)	(25.07)	(26.10)
	N	183	92
Control	x	67.31*	65.92
	(SD)	(23.81)	(25.74)
	N	197	104
		5th Grade	8th Grade
Reading Pretests (in z-scores)			
CDDRE	x	-0.08	-0.07
	(SD)	(0.97)	(1.07)
	N	183	92
Control	x	+0.01	+0.13
	(SD)	(0.98)	(0.97)
	N	197	104
		5th Grade	8th Grade
Mathematics Pretests (in z-scores)			
CDDRE	x	-0.05	-0.04
	(SD)	(1.02)	(1.07)
	N	175	97
Control	x	+0.05	+0.03
	(SD)	(0.92)	(0.90)
	N	190	111

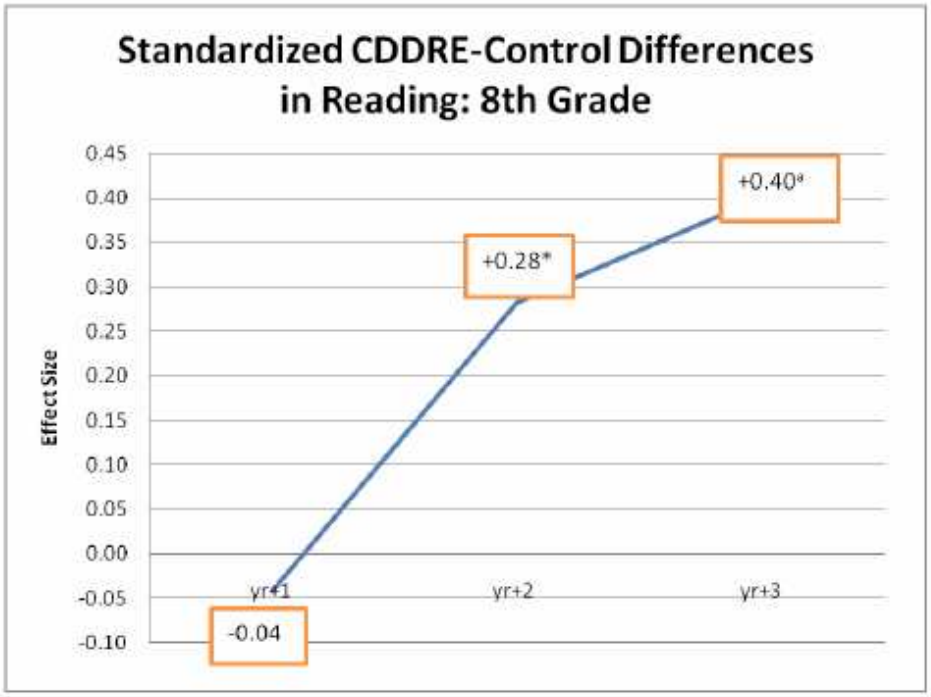
Figure 1



*p<0.05 p**<0.01

	Adjusted posttest scores			
	Baseline	yr+1	yr+2	yr+3
Treatment	-0.08	-0.11	-0.050	0.190
Control	0.01	0.110	0.050	-0.170
SD	0.98	0.93	0.95	0.90
ES	-0.10	-0.24	-0.11	0.40
p-value	0.32	0.01	0.26	0.01
N (schools)	T=183, C=197	T=183, C=197	T=118, C=124	T=58, C=66

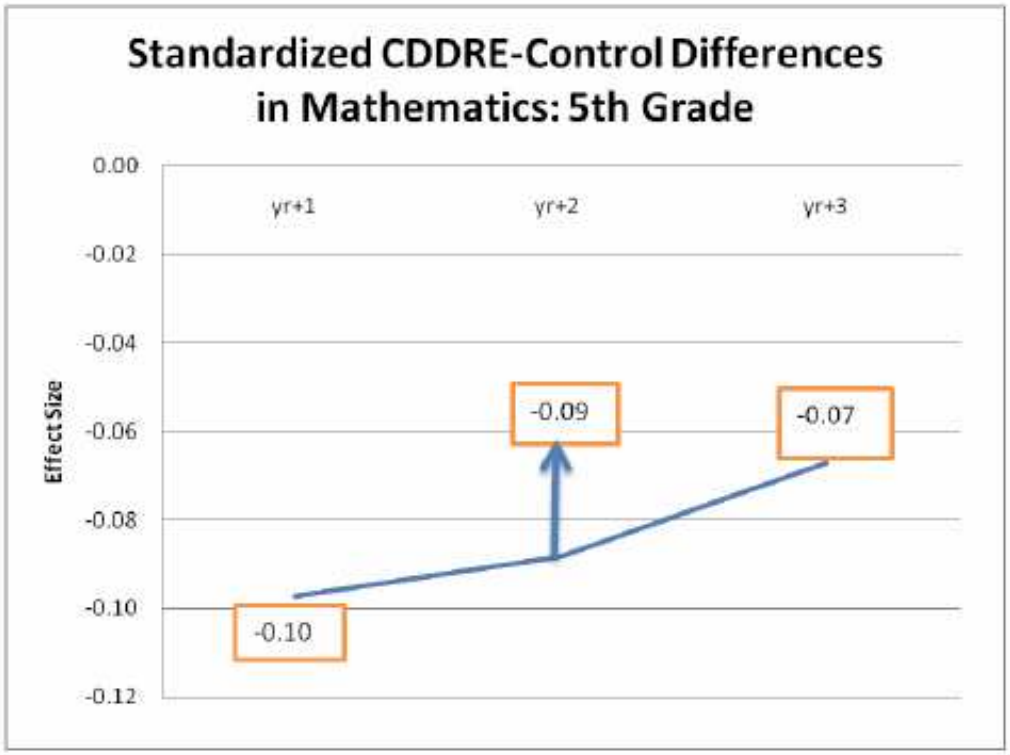
Figure 2



a p<0.10, *p<0.05, and **p<0.01

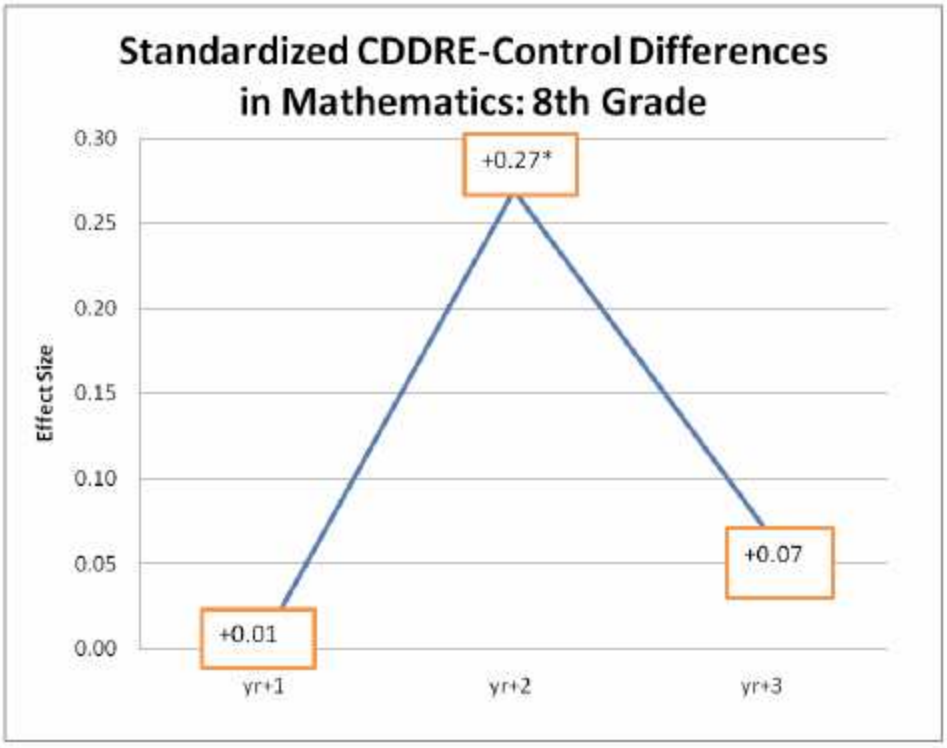
	Adjusted posttest scores			
	Baseline	yr+1	yr+2	yr+3
Treatment	-0.07	-0.02	0.15	0.170
Control	0.13	0.022	-0.15	-0.190
SD	0.97	0.99	1.03	0.90
ES	-0.21	-0.04	0.28	0.40
p-value	0.18	0.64	0.03	0.07
N (schools)	T=92, C=104	T=92, C=104	T=47, C=47	T=26, C=24

Figure 3



	Adjusted posttest scores			
	Baseline	yr+1	yr+2	yr+3
Treatment	-0.05	-0.05	-0.044	-0.030
Control	0.05	0.05	0.039	0.029
SD	0.922	1.03	0.94	0.88
ES	-0.11	-0.10	-0.09	-0.07
p-value	0.96	0.22	0.39	0.66
N (schools)	T=175, C=190	T=175, C=190	T=112, C=127	T=56, C=61

Figure 4



*p<0.05

	Adjusted posttest scores			
	Baseline	yr+1	yr+2	yr+3
Treatment	-0.04	0.006	0.193	0.039
Control	0.03	0.001	-0.033	-0.026
SD	0.90	0.93	0.84	0.90
ES	-0.08	0.01	0.27	0.07
p-value	0.56	0.95	0.05	0.74
N (schools)	T=97, C=111	T=97, C=111	T=51, C=66	T=24, C=32