

The Impact of Comprehensive School Reform
with NSF-Supported Mathematics Curricula
on Urban Middle Grades Student Mathematics Achievement

Martha Abele Mac Iver and Douglas J. Mac Iver

Johns Hopkins University

Paper presented at the Annual Meeting of the American Educational Research Association

April 10, 2007

Chicago

(Submitted for publication)

The authors would like to thank Vaughan Byrnes, Allen Ruby, and Robert Balfanz for their help and comments on this article. The research reported here was supported by the Research on Learning and Education (ROLE) Program at the National Science Foundation, grant number 0411796. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

Recognizing the need to implement standards-based instructional materials with schoolwide coherence led some Philadelphia schools to adopt whole school reform (WSR) models during the late 1990s. This article reports the achievement effects on the Pennsylvania System of School Assessment (PSSA) of the number of years schools had been implementing either a WSR model with NSF-supported math curriculum, or a WSR model without a mathematics curriculum component, from 1997-2000. As hypothesized, math achievement gains (fifth to eighth grade) were positively related to the number of years those schools had implemented a specific math curricular reform. Additional analyses using the PSSA math content clusters and growth between seventh and eighth grades in math scores on the Stanford Achievement Test (SAT9) indicated that the effectiveness of whole school reform with NSF-supported math curricula is broad-based, ranging from helping students to make gains in computation skills to deepening their understanding of higher-order math concepts.

The Impact of Comprehensive School Reform
with NSF-Supported Mathematics Curricula
on Urban Middle Grades Student Mathematics Achievement

Despite encouraging mathematics achievement gains for American adolescents in the Trends in International Mathematics and Science Study (TIMSS) and National Assessment of Educational Progress (NAEP) assessments overall (NCES, 2004; U.S. Department of Education, 2003; O’Sullivan, et al., 2003; Braswell, et al., 2001), urban students remain left behind in mastering the skills they need for successful entry into the American economy. NAEP results in the trial urban assessment (NCES, 2004) indicate that half of public school eighth graders in large central city areas score below basic in mathematics (compared to a third nationwide). Two-thirds or more score below basic in Atlanta, Los Angeles, and the District of Columbia. A gap of 30 to 40 percentage points exists between these students and their more advantaged counterparts even with the most recent gains for all students in the 2003 NAEP for mathematics. These results illustrate the findings from numerous studies showing the unequal opportunities to learn (Blank & Langesen, 1999; Campbell & Silver, 1999; Raudenbush, Fotiu, & Cheong, 1998) and the fewer supports and resources for students in these schools (Flanagan & Grissmer, 2002; Mullis et al., 2000), including differences in teacher quality. Teachers at high-poverty secondary schools are generally less prepared for their positions (NCES, 1999) and, as a result, provide less capable instruction (Grossman, 1990).

Reform efforts aimed at improving urban student achievement in mathematics over the past two decades have involved both district-centered and school-centered approaches. They have also been influenced by the “math wars” among policymakers and other stakeholders as well as by debates over what constitutes compelling evidence of effectiveness (e.g., Battista,

1999; Schoen, Fey, Hirsch, & Coxford, 1999; Cossey, 1999; Bishop, 1999; Grossen & Ewing, 1994; Hiebert, 1999; Loveless, 1997; Carnine & Gersten, 2000; Schoen, 1999; Huntley et al., 2000; Fuson, Carroll, & Drucek, 2000; O'Brien, 1999; Prawat, 1997). Although debate continues to rage over particular means of reforming mathematics education, there remains considerable consensus about the need to implement standards-based curriculum, instructional materials and assessment tools (Jackson & Davis, 2000; RAND Mathematics Study Panel, 2003; Wheelock, 1998), and for schoolwide consistency and coherence in curriculum and instruction rather than the hodgepodge of materials both within the same grade and across grades found in many schools (Balfanz & Mac Iver, 2000; Balfanz, Mac Iver, & Byrnes, 2006; Newmann, Smith, Allensworth, & Bryk, 2001). The need for “academic press” (rigorous standards and high expectations for academic learning) as well as social support in the middle school setting is well documented (Lee, Smith, Perry, & Smylie, 1999; Mac Iver, Ruby, Balfanz, & Byrnes, 2003; Wilson & Corbett, 2001.)

During the 1990s the National Science Foundation sought to improve mathematics (and science) achievement in urban school districts through its Urban Systemic Initiative (USI), followed by the Urban Systemic Program (USP). In urban districts like Philadelphia, NSF's USI program promoted NCTM standards-based curricula such as Math in Context, Mathscape, Connected Math Project, and the University of Chicago School Mathematics Program (UCSMP)/Everyday Math (many of which had been developed with the support of NSF funding). Mini-grants to schools were available for curriculum materials, and USI also offered schools curriculum-specific professional development. Evaluations of the USI program found evidence of longitudinal gains in overall student achievement in science and mathematics in participating districts, with larger gains for those districts with longer USI participation (Borman,

2005; Kim et al., 2001; Kim, Crasco, Leavitt, & Karantonis, 2002). In Philadelphia, the emphasis on decentralized decision-making during the late 1990s (within schools broken into smaller learning communities) meant that curriculum materials and professional development supported by USI often reached only a portion of teachers and students in any particular school, potentially diluting any “school effects” one might have expected.

Concurrent with the district-focused USI initiatives was a national scaling up of Comprehensive School Reform (CSR) initiatives (Kearns & Anderson, 1996). CSR models differed considerably in their emphasis on curriculum and instruction in general, and even more in their focus on mathematics instruction, though some included specific mathematics curricular components (e.g., Talent Development Middle Grades, Talent Development High School, Direct Instruction, Roots and Wings). Congressional approval of the Comprehensive School Reform Demonstration (CSR/D) program in 1997 spurred further expansion in the number of schools seeking to implement whole-school reform models with track records of success, along with numerous studies of the impact of these programs (e.g., Berends, Kirby, Naftel, & McKelvey, 2001; Nunnery, 1998; Stringfield et al., 1997; Herman et al., 1999; Borman, et al., 2003).

In the current policy context, federal funding has been increasingly slated to support “scientifically based” curricular reform programs. Since 2002 the Department of Education has sponsored a “What Works Clearinghouse” (WWC) designed “to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education” (Department of Education, 2007). The study review standards established by the WWC require at least “quasi-experiments with equivalent groups, with no attrition or disruption problems,” with randomized trials greatly preferred. The WWC website features a review of “curriculum-based interventions for improving mathematics achievement for

middle school students.” The review features just five curricula judged to have studies that meet the WWC standards for evidence: Saxon Math, The Expert Mathematician, Cognitive Tutor, I CAN Learn Mathematics Curriculum, and Connected Mathematics Project. Randomized controlled trials (RCTs) were conducted for the first four of these curricula. There was evidence of significantly greater mathematics achievement growth for Cognitive Tutor (compared with McDougal Littell’s Heath Algebra I) and I CAN Learn Mathematics (compared to a non-specified traditional math curriculum). There was no evidence of statistically higher achievement for the Connected Mathematics Project in the quasi-experimental design studies reviewed. The WCC review did not consider or report on studies of the University of Chicago School Mathematics Project’s Transition Mathematics, even though this was the comparison group curriculum in two of the reviewed studies. In neither study was there a significant difference in achievement effects between the curricula.

The narrow focus of the What Works Clearinghouse report on middle school mathematics demonstrates the need for more research attention to middle grades mathematics. Senk and Thompson (2003) and Confrey and Stohl (2005) provide more systematic and thorough analysis of research on middle grades school mathematics curricula, including NCTM-standards based curricula supported by NSF.

With urban students continuing to fall behind their more advantaged peers in mathematics, despite a federal focus on “what works” and recent attempts at whole district reform (D’Amico, Harwell, Stein, & van den Heuvel, 2001; Darling-Hammond, Hightower, Husbands, LaFors, & Young, 2002; Elmore & Burney, 1997; Hightower, 2002; Hightower, Knapp, Marsh, & McLaughlin, 2002; Snipes, Dolittle, & Herlily, 2002), we believe it is important to investigate more thoroughly the contributions and lessons of comprehensive school

reform. Philadelphia, a district with numerous high-minority, high-poverty schools, offers the opportunity to examine the impacts of whole school reform with coherent mathematics curricula on students' academic performance using longitudinal analysis of individual level student achievement growth in mathematics.

Background on Systemic Educational Reform in Philadelphia

Current educational reform efforts in Philadelphia under CEO Paul Vallas (Useem, 2005; Useem & Balfanz, 2002) combine privatization (assignment of some schools to be managed by private educational management organizations rather than the school district) with comprehensive district reforms (coherent instructional programs and strong instructional guidance for teachers through the system-wide implementation of a completely specified core curriculum, pacing guides and curriculum-based assessments every six weeks). This differs substantially from the preceding Children Achieving systemic reform initiative. Launched by Superintendent David Hornbeck in 1995 with funding from the Annenberg Challenge and local matching funds, the Children Achieving initiative relied on standards, decentralization, and accountability reforms while vigorously pursuing fairer, more adequate funding for the District's schools and mandating that all schools be divided into smaller units (small learning communities, or SLCs) that would provide more personalized environments for teaching and learning. In contrast to the current centralized district management of curriculum and instruction, the decentralization under Children Achieving involved moving decision-making, leadership, and support for curriculum and instruction to geographically based clusters of schools and to the SLCs within each school. The district also encouraged interested schools or clusters to use the freedom of decentralization to adopt a whole school reform model if they thought it would help them meet the standards, and the district helped schools and clusters find new funds and allocate

existing funds to implement such a model (Christman, 2001; Corcoran & Christman, 2002; Foley, 2001; Tighe, Wang, & Foley, 2002).

During the Children Achieving period, a large minority of Philadelphia schools pursued a whole school reform strategy (WSR), often with externally developed comprehensive school reform (CSR) models. CSR models differed dramatically in their approach, with some providing specific curricula and professional development support in mathematics, literacy, science and/or history and others focusing more on school processes. Whole school reform models such as Talent Development Middle Grades (TDMG) (Balfanz, Mac Iver, & Byrnes, 2006; Balfanz, Mac Iver & Ryan, 2002; Mac Iver, Ruby, Balfanz, & Byrnes, 2003; Ruby, 2006; Mac Iver, Mac Iver, Balfanz, Plank, & Ruby, 2000; Balfanz, Ruby, & Mac Iver, 2002) and Coalition of Essential Schools (CES) (McDonald et al., 1999) illustrate this contrast in approach. While TDMG made clear recommendations regarding instructional programs and practices and provided sustained curriculum-specific professional development and in-classroom coaching, CES focused more on creating and sustaining professional study groups for administrators and teachers, sharing a set of principles for school reform, and creating a network to equip and support teachers and administrators in their change efforts.

Hypotheses

If comprehensive school reform is important for improving student achievement, one would expect that during the Children Achieving era, when the decentralization policy allowed schools to decide for themselves on curriculum and instructional strategies, students at middle grades schools with sustained implementation of a comprehensive school reform model would show greater achievement gains than students at other schools. And if instructional coherence is particularly important, we would expect that students at schools with a WSR strategy that

emphasized the implementation of a coherent math curriculum would show greater math achievement gains than students at both other (non-math-focused) CSR schools and schools with no specific whole school reform strategy. We hypothesize that students at schools that have implemented a NSF-supported mathematics curriculum school wide will show greater achievement growth than students at other schools because they have received a coherent, standards-based program of instruction from teachers who can interact with each other regarding implementation of the math curriculum used in all classrooms of the school. Focused professional development, shown to have a significant impact on teacher practice (e.g., Desimone, Porter, Garet, Yoon & Birman, 2002), is likely to be much more prevalent at schools using the same curriculum school-wide than at schools with a jumble of curriculum. While many Philadelphia schools used Urban Systemic Initiative (USI) funds to purchase standards-based mathematics curricula, these materials were often used in only some classrooms at only some grade levels. At these schools that lacked a mathematics-related whole school reform strategy, students sometimes received some standards based instruction, but often encountered an uneven and changing curriculum from grade to grade with teachers who had fewer opportunities to learn from and improve each other's math teaching. Thus, we expected that students at schools with a mathematics-related whole school reform strategy that included implementation of a coherent, NSF-supported curriculum in all classrooms schoolwide would make greater mathematics achievement gains than students in other schools.

Methodology

Data and Measures

Data on the primary independent variable of interest in this study, whether Philadelphia schools were implementing a mathematics-oriented whole school reform (Math WSR) strategy or other comprehensive school reform (Other CSR) model during the late 1990s, were obtained from archival records maintained by Research for Action. These data, based on Philadelphia district records (supplemented by records obtained from CSR model development teams) included information on which particular whole school reform strategies were in place at the Philadelphia schools that served eighth graders during this period, and how long they had been implemented.

Among the 86 regular schools with eighth grade students there were 12 Math WSR schools in 1999-2000 (the outcome data year). These schools implemented schoolwide one of the math programs classified by Confrey and Stohl (2004) and Senk and Thompson (2003) as an “NSF-supported” curriculum or, in the case of the University of Chicago School Mathematics Project’s late elementary and early secondary components, as a partly NSF-supported/ partly commercially generated “hybrid” curriculum: Connected Mathematics, Math in Context (MIC), Mathscape, Talent Development Mathematics Program (TDMP), and University of Chicago School Mathematics Project (UCSMP). The adoption of one of these evidence-based math programs schoolwide was largely in response to NSF’s USI program in Philadelphia, which promoted these curricula and offered schools curriculum-specific professional development in their use. (Because curriculum selection was so decentralized, even within particular schools, in Philadelphia at this time, USI funding often supported implementation of the NSF-supported mathematics curriculum in only certain classrooms within a school, and data were not available

to link students to a particular curriculum at this level.) TDMP and UCSMP are listed separately above even though TDMP typically involves implementation of UCSMP materials – *Everyday Mathematics* in grades 5 and 6 and *Transition Mathematics* and *Algebra* in grades 7 and 8. (Although TDMG recommends Chicago materials most highly, it supports schools that choose to implement one of the other NSF-supported mathematics curricula schoolwide instead.) Schools that embraced one of these curricula as a key piece of their whole school reform efforts typically drew upon more than one funding stream (such as CSR grants, regular curriculum budgets, schoolwide Title 1 funds, and USI mini-grants) to purchase them. Only schools using the NSF-supported curricula school wide were coded as Math WSR sites.

There were 28 “Other CSR” schools in 1999-2000 that implemented a non-math-focused CSR model as their whole school reform strategy. These included the following models and programs: ATLAS Communities, Audrey Cohen, Co-Nect, Coalition of Essential Schools, Community of Learning, Core Knowledge, First Steps, Modern Red Schoolhouse, Reading Recovery (probably not in middle grades at the K-8 school), Talent Development Literacy Program, and SFA/Roots & Wings (probably not in middle grades at the K-8 school). The list of “Other CSR” models reflects several waves of school reform. The Coalition of Essential Schools began in the mid-1980s with a group of schools committed to implementing the principles articulated bySizer in *Horace’s Compromise*. Its adoption by Philadelphia schools occurred in the mid 1990s. Several of the CSR models (ATLAS, Audrey Cohen, Co-Nect, Modern Red Schoolhouse, and Roots and Wings) were those supported by the New American Schools Corporation during the early 1990s, and adopted somewhat later by Philadelphia schools. Congressionally appropriated funding to support implementation of comprehensive

school reform designs (CSR grants) began in 1998, and many schools selected a “brand name” CSR model or proposed their own WSR strategy to qualify for the three-year awards.

The remaining 46 schools were coded as “No CSR” sites because district records indicated that they did not have a CSR model or articulated WSR strategy at that time. Unfortunately, no other data were available regarding what curriculum or reform efforts were in place at these schools. We also created school level variables indicating the number of years each school had been implementing a Math WSR or Other CSR (using data from RFA records, supplemented with data supplied by reform developers). This variable ranged from 0 to 3 for Math WSR models, and from 0 to 7 for Other CSR models.

Table 1 summarizes average characteristics of the three types of schools (Math WSR sites, other CSR sites, and other schools). The Math WSR schools tended to have greater numbers of eighth grade students (since there were proportionately more middle schools in this group). Other CSR schools had significantly higher attendance rates than other types of schools. The data indicate that Math WSR schools did not have significantly fewer challenges than other types of schools.

(Table 1 about here)

Other school level measures included school type (1=K-8 school, 0 = middle school) and a measure of the average achievement level of the school’s entering sixth grade student – the school-wide average fifth grade spring mathematics score on the Pennsylvania System of School Assessment (PSSA). At the individual level, measures were available for gender (female=1), ethnicity group, special education status, and English language learner status. In addition, students were coded for whether they spent all three years (6th, 7th, and 8th grade) in the same school or not, since mobility has been shown to have a negative impact on student achievement.

Our dependent variable measure is achievement growth on the Pennsylvania System of School Assessment (PSSA), “a standards based criterion-referenced assessment used to measure a student's attainment of the academic standards while also determining the degree to which school programs enable students to attain proficiency of the standards” (Pennsylvania Department of Education, 2007). PSSA scores have been found to be highly correlated with scores on nationally normed tests such as the CTBS/Terra Nova (Human Resources Research Organization, 2004). Prior to 2004, when NCLB mandates led to more extensive testing, the PSSA reading and mathematics tests were administered to every Pennsylvania student in grades 3, 5, 8 and 11. The test is vertically equated so that it is possible to measure scale score growth over time (in different grades). In these analyses we use data for the cohort of students who traversed fifth through eighth grade in 1997-2000 (the key years of the Children Achieving initiative under Superintendent Hornbeck). The analyses are based on scores for 9,421 students in 86 regular schools.ⁱ “Content cluster” scores (raw number correct) on the PSSA mathematics test were also available for Spring 2000. The content clusters for the eighth grade assessment include: Numbers, Number Systems, and Number Relationships; Computation and Estimation; Measurement and Estimation; Mathematical Reasoning and Connections; Mathematical Problem- Solving and Communication; Statistics and Data Analysis; Probability and Predictions; Algebra and Functions; Geometry; Trigonometry; and Concepts of Calculus.

Modeling Student Achievement Growth

We use multilevel change models (Raudenbush & Bryk, 2002; Seltzer, Choi, & Thum, 2003) to estimate the impact of the comprehensive school reforms on students' achievement growth during the middle grades, because students are nested within schools. With just two time points per student on the state tests, we model initial status and total growth rather than a full

“growth curve.” In estimating 3-level growth models, we specify a within-student model, a between-student model, and a school level model.

At level 1, we model students’ achievement growth (in NCE scores) as a function of grade (a dummy variable coded “0” if the score is from a student’s fifth-grade year, and a “1” if the score is from a student’s eighth-grade year.) Thus, the coefficient for the intercept (P0) represents students’ prior achievement in the spring of fifth grade and the slope coefficient for grade (P1) represents students’ cumulative achievement growth between the spring of fifth and the spring of eighth grade. The simple within-student model is: $Y = P0 + P1(\text{EIGHTH}) + E$.

At level 2, the between-student model, we take account of differences between students in prior achievement (at the end of the elementary grades) and in achievement growth (during the middle grades) that are associated with student characteristics (e.g., gender, ethnicity, and special education status). Specifically, the model for prior math achievement takes account of the lower prior achievement observed for students who attended low-income schools as fifth graders, and for English language learners, special education students, and females. It also takes account of the higher prior achievement observed for Asians, Hispanics, and Caucasians than for African-Americans. The model for math achievement growth takes account of the higher growth observed for Asians than for other ethnic groups, and the higher growth shown by females and students who attended the same school throughout the middle grades. Therefore, our between-student model is:

$$\text{Prior Achievement: } P0 = B00 + B01(\text{SPECIAL EDUCATION}) + B02(\text{FEMALE}) + B03(\text{ASIAN}) + B04(\text{HISPANIC}) + B05(\text{CAUCASIAN}) + B06(\text{ENGLISH LANGUAGE LEARNER}) + R0$$

$$\text{Achievement Growth: } P1 = B10 + B11(\text{FEMALE}) + B12(\text{ASIAN}) + B13(\text{SAMESCHOOL})$$

At Level 3, the school level, we estimate the impact of school characteristics (K8 or middle school, and average achievement level of entering students) and the years of reform interventions (YRS OF MATH WSR and YRS OF OTHER CSR) on B10, students' adjusted mean mathematics achievement growth. Since average starting achievement level of students in a school was highly correlated with that school's percentage of low-income students, the latter variable was not included in the model for achievement growth to minimize multicollinearity issues. Two other available school level variables -- percentage of teachers with certification and teachers' average years of experience in the district -- were also highly correlated with average achievement level of entering students. Because of this, there were no consistent significant effects of these variables in preliminary models, and they were excluded from the final models. Preliminary analyses indicated that the residual parameter variance for all coefficients except B00, B01 and B04 should be set to zero. Therefore, our level 3 model was:

$$B00 = G000 + U00$$

$$B01 = G010 + U01$$

$$B02 = G020$$

$$B03 = G030$$

$$B04 = G040 + U04$$

$$B05 = G050$$

$$B06 = G060$$

$$B07 = G070$$

$$B10 = G100 + G101(\text{AVERAGE } 5^{\text{TH}} \text{ GRADE MATH SCORE OF SCHL'S STUDENTS}) + \\ G102 (\text{YRS OF MATH WSR}) + G103(\text{YEARS OF OTHER CSR}) + G104(\text{K8}) + U10$$

$$B11 = G110$$

$$B12 = G120$$

$$B13 = G130$$

Results

Table 2 shows the HLM estimates for the Spring 2000 cohort of eighth-graders. Not surprisingly, demographic variables were significant predictors of both students' baseline (fifth grade) achievement and their achievement growth during the middle grades. Further, as we hypothesized, the math achievement gains displayed by students between the fifth grade and eighth grade were positively related to the number of years those schools had implemented a specific math curricular reform. On average, students at these schools outgained students at non-implementing schools by 1.12 NCE points for each year the school had implemented the reform (or more than 3 NCE points for schools that had implemented the reform for three years). Restating this effect of Math WSR as a three-year effect size, students gained 0.19 standard deviations more in math achievement at schools that implemented a Math WSR throughout the students' three years in the middle grades than did students in schools that did not implement a math WSR. (The standard deviations in this sample of the fifth grade and eighth grade NCE scores were 18.0 and 17.7, respectively.) Gains at schools that implemented other CSR models were not significantly larger than those at schools without a reform model in place. Gains were significantly higher at K8 schools (compared to middle schools), controlling for the schools' average incoming (fifth grade) scores. The significant negative coefficient for schools' average

incoming score (the average starting math level for students in the school) indicates that growth was higher at schools that had lower starting points.

(Table 2 about here)

In addition, we conducted analyses using each of the PSSA math content cluster scores to determine in which particular math domains the effect of the math WSR implementation was most pronounced. Because previous years' content cluster scores were not available, a two-level model was more appropriate than the 3-level model used for the overall PSSA analyses. In estimating a two-level model, we specify a between-student model and a school level model. At level 1 (the between-student model), we model students' eighth grade content cluster scores as a function of demographic variables and prior fifth grade PSSA overall math score. So our between-student model is: $Y = B00 + B01(5^{\text{th}} \text{ GRADE MATH SCORE}) + B02(\text{SPECIAL EDUCATION}) + B03(\text{FEMALE}) + B04(\text{ASIAN}) + B05(\text{HISPANIC}) + B06(\text{CAUCASIAN}) + B07(\text{ENGLISH LANGUAGE LEARNER}) + R0$.

At Level 2, the school level, we estimate the impact of the years of reform interventions (YRS OF MATH WSR and YRS OF OTHER CSR) as well as school characteristics (whether the school is a K-8 or middle school, how well, on average, the schools' students were performing in mathematics on their entry to the middle grades school) on students' content cluster score:

$$B00 = G100 + G101(\text{K8}) + G102(\text{AVERAGE } 5^{\text{TH}} \text{ GRADE MATH SCORE OF SCHL'S STUDENTS}) + G103(\text{YRS OF MATH WSR}) + G104(\text{YEARS OF OTHER CSR}) + U00$$

Analyses were conducted using data from 9,280 students at 86 schools. Preliminary analyses indicated that the residual parameter variance for all level 1 coefficients should be set to zero.

Table 3 summarizes the coefficients for YEARS OF MATH WSR for each of the content clusters, as well as the three-year effect size (in bold) for each one. All the coefficients were positive, and the effect of YEARS OF MATH WSR was significantly positive ($p \leq .05$) for the following math content areas: 1) geometry, 2) concepts of calculus, 3) computation and estimation; and 4) measurement. In addition, the effect of YEARS OF MATH WSR was positive at $p \leq .10$ for numbers, number systems and number relationships, probability and predictions, and statistics. These results suggest that the effectiveness of whole school reform with NSF-supported math curricula is broad-based, ranging from helping students to make gains in computation skills to deepening their understanding of higher-order math concepts.

(Table 3 about here)

An alternative measure of students' mathematics achievement growth during their final year in the middle grades was also available. Analyses of growth between seventh and eighth grades in math scores on the Stanford Achievement Test (SAT9) for the same cohort of students (eighth graders in 1999-2000) offer additional evidence of the positive influence of attending a school that was implementing a whole school reform with a particular math curriculum focus. Figure 1 shows the differences in final-year NCE growth for students at schools with math-specific whole school reform models, other whole school reform models, and no whole school reform models. As the figure indicates, student growth at schools with math reforms was particularly pronounced in math problem solving. Students at these schools also had gains in math computation, while other students lost ground in this area between seventh and eighth grades. HLM analyses of the impact of YEARS OF MATH WSR yielded a significantly positive coefficient (1.2 NCEs, $p=.043$) for YEARS of MATH WSR on eighth grade math problem-solving NCE score (controlling for fifth grade math score and other demographic

variables) in the sample of 5,975 students at the 50 high-poverty (80% or more low-income students) schools (Table 4).¹ The three-year effect size of Math WSR on this mathematics problem-solving subtest was 0.24 standard deviations.

(Figure 1 and Table 4 about here)

Discussion

These findings provide evidence across multiple tests and subtests that sustained implementation of mathematics-oriented whole school reforms makes a difference in student mathematics achievement. The impact was not based simply on whether or not the school implemented a WSR strategy using a standards-based math curriculum, but on the number of years the strategy or model had been implemented at the student's eighth grade school. This indicates a stronger pattern of relationship between reform implementation and student achievement. The effect remained even when direct measures of a school's capacity and student population (percentage of non-certified teachers, teachers' average years in the district, or students' average incoming achievement level) were included in the model. A particularly important finding in this study of Philadelphia's middle grades students is the fact that exposure to these innovative and standards-based math curricula, so often criticized by traditionalists as "fuzzy math" and even accused of being detrimental to students' ability to perform well in higher mathematics, actually had a significantly positive effect on student achievement on the PSSA in computation as well as on a wide variety of higher-level concepts.

The available data do not allow us to distinguish between different types of coherent mathematics curricula. All the schools implementing a mathematics-oriented school reform during this period in Philadelphia were using a NCTM standards-based curriculum (developed with funding support from the National Science Foundation). No Philadelphia schools were on

¹ Results were not significant in the larger sample of schools with more diversity in poverty status.

record as using such computation-focused programs as Direct Instruction Mathematics or Saxon Math as a whole school math reform effort in the middle grades, so we cannot compare the impact of two different types of curricula. Given other research findings (e.g., Mac Iver, Kemper, & Stringfield, 2003; Mac Iver, 2005), we would expect that there could have been significant gains in computation using curricula such as DI Math, but probably not in more conceptual areas of mathematics.

This study of the impact of WSR in Philadelphia has other limitations common to most longitudinal studies of archival student level and school level data. We were limited to the variables collected by the district during the particular period. We have no data regarding the quality of implementation of the various reform efforts in schools and no systematic information about the math curricula used by schools that were not implementing a whole school math reform strategy. Similar analyses of student achievement growth in the cohorts reaching eighth grade during the tumultuous transition period following the end of Children Achieving (2001 and 2002) did not yield a significant effect for years of whole school math reform, suggesting that the district turmoil and budget cutbacks during this period quickly rendered the positive effects reported here unsustainable. Another reason for the diminished effects in these two later cohorts may be that the district appears not to have kept school level data on implementation of reform models for these transition-era cohorts because of the considerable uncertainty and instability during those years. Nonetheless, this analysis of comprehensive and whole school reform in one district over several years makes an important contribution to the growing literature regarding the impact of such reform efforts because of the sheer number of schools involved and our ability to distinguish between reform efforts that emphasized a coherent math curriculum and those that did not.

We also contend that this study of whole school reform is an important contribution to the research community's current focus on whole district reform. We interpret Philadelphia's decision in 2003 to mandate a district-wide implementation of an NSF-supported curriculum (Math in Context -- with pacing guides, regular benchmark assessments, and targeted professional development for teachers) as an expansion of the principles underlying the math-focused whole school reform efforts of some schools under the Children Achieving initiative. In many ways, that initiative represented a district-wide scaling up of comprehensive school reform, with professional development and coaching focused on single standards-based curriculum, providing a common focus for teachers and a coherent curricular pathway for even mobile students who moved among district schools. Elsewhere (Mac Iver & Mac Iver, 2006) we discuss the dramatic gains in math achievement systemwide from 2003 to 2004, the first year of this district-wide implementation of a standards-based math curriculum undergirded by extensive professional development for teachers. In on-going research, we are analyzing the effect of a shift to a district-mandated mathematics instructional program on schools that were doing well with a comprehensive school reform model under the previous policy environment. To what extent did the district-wide "scaling up" either contribute to further gains at these schools, or upset a particular reform strategy already well in place? We are also exploring the continued process of implementing a district-wide curriculum in the context of large-scale teacher turnover. To what extent does the district continue to provide the undergirding professional development support for implementing the mathematics curriculum, and to what extent does this affect student achievement gains? Our future research will address the impact of variation in implementation, professional development support, and teacher quality variables on student outcomes when all district schools are using the same standards-based mathematics instructional program.

In an era of emphasis on district-wide reforms, districts will increasingly be faced with the issue of how to treat successful pockets of reform (often based on CSR models) that may be marching to a different drummer than the envisioned district-wide model. (The experience of Baltimore with Direct Instruction is a particular example of this issue. See Mac Iver, 2004.) It is important in this context to remember the underlying relationship between the principles of comprehensive school reform and the goals of a district-wide effort to implement coherent, standards-based instruction and support for teachers to deliver it. A crucial question for researchers to continue to explore now is the impact of the “support delivery system” to schools seeking to provide high quality standards-based curriculum to students in every classroom, every day. To what extent do districts require external help (e.g., from CSR developer teams, from EMOs or other external actors) to deliver the support to teachers required to implement high quality instruction? Is district-wide consistency in curriculum and professional development essential, or can a diversity of reform efforts coexist within the same district framework (and under what conditions)? The evidence that implementation of coherent, standards-based mathematics instructional programs helps to improve student achievement is clear. How to scale this up more effectively so that all students leave school with the skills they need for productive work lives is the question that remains before us.

References

- Balfanz, R., & Mac Iver, D. J. (2000). Transforming high-poverty urban middle schools into strong learning institutions: Lessons from the first five years of the Talent Development Middle Grades. *Journal of Education for Students Placed at Risk*, 5, 137-158.
- Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). *The implementation and impact of evidence based mathematics reforms in high poverty middle schools: A multi-site, multi-year study*. *Journal for Research in Mathematics Education* 37, 33-64.
- Balfanz, R., Mac Iver, D.J., & Ryan, D. (2002). Enabling “algebra for all” with a facilitated instructional program: A case study of a Talent Development Middle Grades. In V. Anfara, (Ed.), *Middle school curriculum, instruction, and assessment: Volume 2 in the handbook of research in middle level education*. Greenwich, CT: Information Age Publishing.
- Balfanz, R., Ruby, A., & Mac Iver, D. (2002). Essential components and next steps for comprehensive whole-school reform in high poverty middle schools. In S. Stringfield & D. Land (Eds.), *Educating At-Risk Students: One Hundred-first Yearbook of the National Society for the Study of Education, Part II*. (pp. 128-147). Chicago, IL: NSSE.
- Battista, M.T. (1999, February). The mathematical miseducation of America’s youth. *Phi Delta Kappan*, 425-433.
- Berends, M., Kirby, S. N., Naftel, S., & McKelvy, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Washington, DC: Rand Corporation.
- Blank, R., & Langesen, D. (1999). *State indicators of science and mathematics education*. Washington, DC: Council of Chief State School Officers.

- Borman, G., Hewes, G., Rahuba, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research* 73, 125-230.
- Borman, et. al. (2005). *Meaningful urban education reform: Confronting the learning crisis in mathematics and science*. Albany: State University of New York Press.
- Braswell, J.S., Lutkus, A.D., Grigg, W.S., Santapau, S.L., Tay-Lim, B. and M. Johnson (2001). *The nation's report card: Math 2000*. NCES 2003-453. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Washington, DC.
- Campbell, P.F., & Silver, E. A. (1999). *Teaching and learning mathematics in poor communities*. Reston, VA: National Council of Teachers of Mathematics.
- Carnine, D. & Gersten, R. (2000). The nature and roles of research in improving achievement in mathematics. *Journal for Research in Mathematics Education*, 31(2), 138-43.
- Christman, J.B. (2001, December). *Powerful ideas, modest gains: Five years of systemic reform in Philadelphia middle schools*. Philadelphia, PA: Consortium for Policy Research in Education.
- Corcoran, T., & Christman, J.B. (2002, November). *The limits and contradictions of systemic reform: The Philadelphia story*. Philadelphia, PA: Consortium for Policy Research in Education.
- Cossey, R. (1999, February). Are California's math standards up to the challenge? *Phi Delta Kappan*, 441-443.
- D'Amico, L., Harwell, M., Stein, M.K., & van den Heuvel, J. (2001). *Examining the implementation and effectiveness of a district-wide instructional improvement effort*.

Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, Washington.

Darling-Hammond, L., Hightower, A.M., Husbands, J.L., LaFors, J.R., & Young, V.M. (2002, April). *Building instructional quality: Inside-out, bottom-up, and top-down perspectives on San Diego's school reform*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Desimone, L.M., Porter, A.C., Garet, M.S., Yoon, K.S., & Birman, B.F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.

Elmore, R.F., & Burney, D. (1997). *School variation and systemic instructional improvement in Community School District #2, New York City*. Pittsburgh, PA: University of Pittsburgh, HPLC Project, Learning Research and Development Center.

Flanagan, A. & Grissmer, D. (2002). The role of federal resources in closing the achievement gap. In J.E. Chubb & T. Loveless (Eds.), *Bridging the achievement gap* (pp. 199-225). Washington, DC: Brookings Institution Press.

Foley, E. (2001, August). *Contradictions and control in systemic reform: The ascendancy of the central office in Philadelphia Schools*. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education, Graduate School of Education.

Fuson, K.C., Carroll, W.M., and Drueck, J.V. (2000). Achievement results for second and third graders using the standards-based curriculum Everyday Mathematics. *Journal for Research in Mathematics Education*, 31(3), 277-295.

Grossen, B. & Ewing, S. (1994). Raising mathematics problem-solving performance: Do the NCTM teaching standards help? *Effective School Practices*, Spring 1994.

Grossman, P.L. (1990). *The making of a teacher: Teacher knowledge and teacher education.*

New York: Teachers College Press.

Herlihy, C. & Kemple, J. (2003, April). *Impacts of the Talent Development Middle Grades*

Model on achievement, attendance, and promotion: A third-party evaluation. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Herman, R., Aladjem, D., McMahon, P., Massem, E., Mulligan, I., O'Malley, A., Quinones, S.,

Reeve, A., & Woodruff, D. (1999). *An educator's guide to schoolwide reform.*

Washington, DC: American Institutes for Research.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for*

Research in Mathematics Education, 30(1), 3-19.

Hightower, A. (2002). *San Diego's big boom: District bureaucracy supports culture of*

learning. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Hightower, A., Knapp, M. Marsh, J., & McLaughlin, M. (2002). *School districts and*

instructional renewal. New York: Teachers College Press.

Huntley, M.A., Rasmussen, C., Villarubi, R., Santong, J., Fey, J. (2000). Effects of standards-

based mathematics education: A study of the Core-Plus Mathematics project algebra and functions strand. *Journal for Research in Mathematics Education*, 31(3), 328-361.

Jackson, A.W., & Davis, G.A. (2000). *Turning Points 2000: Educating adolescents in the 21st*

century. New York: Teachers College Press.

Kearnes, D. & Anderson, J. (1996). Sharing the vision: Creating New American Schools. In

Stringfield, S., Ross, S., & Smith, L. (Eds.) *Bold plans for school restructuring* (pp. 9-23). Mahwah, NJ: Erlbaum..

Kim, J.J., Crasco, L.M., Smith, R. B., Johnson, G., Karantonis, A., & Leavitt, D.J. (2001).

Academic excellence for all urban students: Their accomplishment in science and mathematics. Norwood, MA: Systemic Research, Inc.

Kim, J.J., Crasco, L.M., Leavitt, D.J., & Karantonis, A. (2002). *Urban school key indicators of science and mathematics education: 2001.* Vol. II: Cohort 94 school districts. Norwood, MA: Systemic Research, Inc.

Lee, V.E., Smith, J.B., Perry, T.E., & Smylie, M.A. (1999). *Social support, academic press, and student achievement: A view from the middle grades in Chicago.* Chicago, IL: Consortium on Chicago School Research.

Loveless, T. (1997). The second great math rebellion. *Education Week* (10/15/97).

Mac Iver, D., Mac Iver, M., Balfanz, R., Plank, S., & Ruby, A. (2000). Talent Development Middle Grades: Blueprint and results for a comprehensive whole-school reform model. In M. Sanders (Ed.), *Schooling Students Placed at Risk: Research, Policy, and Practice in the Education of Poor and Minority Adolescents* (pp. 261-288). Mahwah, NJ: Erlbaum.

Mac Iver, D.J., Ruby, A., Balfanz, R. & Byrnes, V. (2003). Removed from the list: A comparative longitudinal case study of a reconstitution-eligible school. *Journal of Curriculum and Supervision*, 18 (3): 259-289.

Mac Iver, M. & Mac Iver, D.J. (2006). Which bets paid off? Early findings on the impact of private management and K-8 conversion reforms on the achievement of Philadelphia students. *Review of Policy Research* 25(5), 1077-1093.

- Mac Iver, M. (2005). Getting a running start on reading and math? A four-year cohort study of the achievement impact of Direct Instruction. Unpublished paper.
- Mac Iver, M. (2004). Systemic supports for comprehensive school reform: The “institutionalization” of Direct Instruction in an urban school system. *Journal of Education for Students Placed at Risk* 9, 303-321.
- Mac Iver, M., Kemper, E., & Stringfield, S (2003). *The Baltimore Curriculum Project: Final report of the four-year evaluation study*. CRESPAR Report #62. Baltimore: Center for Social Organization of Schools, Johns Hopkins University.
- McDonald, J., Hatch, T., Kirby, E., Ames, N., Haynes, N., & Joyner, E. (1999). The Coalition of Essential Schools in its second decade. In J. McDonald, T. Hatch, E. Kirby, N. Ames, N. Haynes & E. Joyner (Eds.), *School reform behind the scenes: How ATLAS is shaping the future of education* (pp. 45-66). New York: Teachers College Press.
- Mullis, I. V. S, Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O’Connor, K. M., Chrostowski, S. J. & Smith, T. A. (2000, December) *TIMSS 1999 International Mathematics Report: Findings from IEA’s Repeat of the Third International Mathematics and Science Study at Eighth Grade*. Boston, MA: International Study Center, Lynch School of Education.
- National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education. (1999). *Teacher quality: A report on the preparation and qualifications of public school teachers*. Washington, DC: U.S. Department of Education.
- National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee,

- Center for Education, Division of Behavioral and Social Sciences and Education.
Washington, DC: National Academy Press.
- Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A.S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23(4), 297-321.
- Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society*, 30 (3), 277-295.
- O'Brien, T.C. (1999). Parrot math. *Phi Delta Kappan*, 434-440.
- O'Sullivan, C., Lauko, M., Grigg, W., Qian, J. and Zhang, J. (January 2003). *The nation's report card: Science 2003*. NCES 2003-453. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Washington, DC.
- Pennsylvania State Department of Education. (2007). Assessment. Retrieved January 25, 2007 from
[//www.pde.state.pa.us/a_and_t/site/default.asp?g=0&a_and_tNav=|630|&k12Nav=|11411](http://www.pde.state.pa.us/a_and_t/site/default.asp?g=0&a_and_tNav=|630|&k12Nav=|11411)
- Prawat, R. (1997). Fuzzy math, new math, and Dewey. *Education Week*, December 10, 1997.
- RAND Mathematics Study Panel (2003). *Mathematical proficiency for all students: toward a strategic research and development program in mathematics education*, MR-1643-OERI, Santa Monica, CA: The RAND Corporation.
- Raudenbush, S.W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Newbury Park: CA: Sage.
- Raudenbush, S.W., Fotiu, R.P., & Cheong, Y.F. (1998). Inequality of access to educational resources: A national report card for eighth grade math. *Educational Evaluation and Policy Analysis*, 20(4), 253-268.

- Ruby, A. (2006). Improving science achievement at high-poverty urban middle schools, *Science Education*, 90, 6.
- Schoen, H. Fey, J. Hirsch, C. & Coxford, A. (1999, February). Issues and options in the math wars. *Phi Delta Kappan*, 444-453.
- Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, 25(3): 263-286.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002, September). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: The Council of the Great City Schools.
- Stringfield, S., Millsap, M., Winfield, L., Brigham, N., Yoder, N., Moss, M., Nesselrodt, P., Schaffer, E., Bedinger, S., & Gamse, B. (1997). *Urban and suburban/rural special strategies for educating disadvantaged children: Second year report*. Washington, DC: U.S. Department of Education.
- Tighe, E., Wang, A., & Foley, E. (2002, February). *An analysis of the effect of Children Achieving on student achievement in Philadelphia elementary schools*. Philadelphia, PA: Consortium for Policy Research in Education.
- U.S. Department of Education. (2007). What Works Clearinghouse overview. Retrieved January 25, 2007 from www.whatworks.ed.gov/whoweare/overview.html.

- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2003). *The Nation's Report Card: Mathematics Highlights 2003*. Author, Washington, DC.
- Useem, E.L. & Balfanz, R. (2002). Comprehensive district reform: Philadelphia's grand experiment. *Benchmarks* 4(1), 1-10.
- Useem, E.L. (2005, August). Learning from Philadelphia's school reform: What do the research findings show so far? Paper presented at the American Sociological Association Sociology of Education Section No Child Left Behind Conference, Philadelphia, PA.
- Wheelock, A. (1998). *Safe to be smart: Building a culture for standards-based reform in the middle grades*. Columbus, OH: National Middle School Association.
- Wilson, B. L. & Corbett, H.D. (2001). *Listening to urban kids: School reform and the teachers they want*. Albany, NY: SUNY Press.

Table 1

Differences between Schools with Whole School Reform Models and Other Schools

School Characteristics	Math WSR (n=12)	Other CSR (n=28)	No CSR (n=46)
Average 5 th grade PSSA NCE Math Score	27.9	33.7	30.7
Student Attendance	87.5	90.5*	88.8
% Low Income Students	78.5	71.5	78.1
% Certified Teachers	93.8	93.4	93.9
Average Years of Teacher Experience	10.2	13.5	12.8
Average Grade 5-8 Enrollment	803	567*	498*

* Significantly different from Math WSR schools at $p < .05$.

Table 2. Modeling Prior Math Achievement and 5th to 8th Grade Achievement Growth on the PSSA Mathematics Test: HLM Estimates

Fixed effect	Coefficient	Robust Std. Error	p-value
Level 1 (within students)			
Intercept	32.4	0.61	.000
Model for P0 (NCE score in Spring of 5 th grade)			
Level 2 (between students)			
%LOW INC (5 th grade school)	- 0.2	0.02	.000
SPECIAL EDUC	- 8.5	1.01	.000
ENG LANG LEARNER	- 7.4	3.71	.044
ASIAN	14.2	1.06	.000
HISPANIC	1.4	0.54	.009
CAUCASIAN	8.0	0.56	.000
FEMALE	- 1.3	0.33	.000
Model for P1 (Growth Between Spring of 5 th Grade and Spring of 8 th Grade)			
Intercept	1.01	0.64	.117
ASIAN	3.04	0.64	.000
FEMALE	1.02	0.29	.001
SAME SCHOOL	1.01	0.46	.029
AVG INCOMING MATH NCE (END OF 5 TH GRADE)	-0.50	0.04	.000
K8	2.79	0.63	.000
YRS OF MATH WSR	1.12	0.34	.002
YRS OF OTHER CSR	0.30	0.21	.159

Table 3. Effect of Years of Whole-School Math Reform on Eighth Grade PSSA Mathematics Content Cluster Scores, Spring 2000

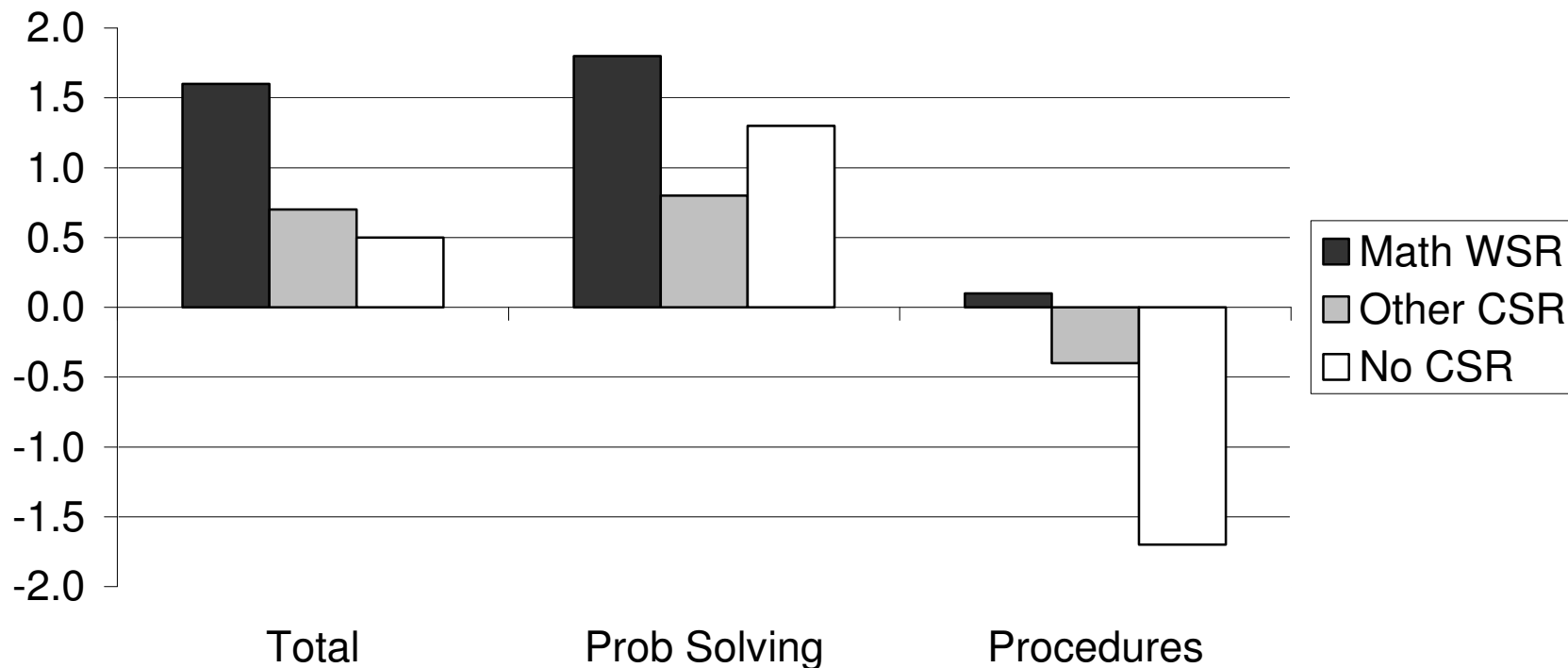
Content Cluster	Items	Std. Dev.	Coefficient	3 year Effect Size	P-Value
Geometry	8	1.8	0.09	0.15	0.001
Calculus	4	1.1	0.05	0.14	0.006
Number Concepts	13	2.6	0.11	0.13	0.066
Probability	4	1.2	0.05	0.13	0.057
Computation	21	4.2	0.16	0.11	0.013
Measurement	12	2.7	0.08	0.09	0.048
Problem Solving	15	3.2	0.09	0.08	0.227
Algebra	8	1.8	0.05	0.08	0.225
Statistics	7	1.7	0.04	0.07	0.104
Reasoning	4	1.1	0.02	0.05	0.312
Trigonometry	4	1.1	0.01	0.03	0.571

Coefficients are based on 2-Level HLM models of 8th grade Content Cluster scores, predicted by student demographic characteristics, fifth grade PSSA mathematics score, and school characteristics besides years of Math-focused whole school reform (school structure, average fifth grade math score of incoming students, years of other comprehensive school reform). Analyses based on 9,280 students at 86 schools.

Table 4. Modeling 8th Grade Achievement at High Poverty Schools on the SAT9 Mathematics Problem Solving Test: HLM Estimates

Fixed effect	Coefficient	Robust Std. Error	p-value
Level 1 (between students)			
Intercept	38.35	0.77	.000
5 th grade math NCE score	0.49	0.02	.000
SPECIAL EDUC	- 2.29	0.87	.011
ASIAN	7.03	1.53	.000
HISPANIC	1.29	0.54	.017
CAUCASIAN	2.73	0.67	.000
Level 2 (between schools)			
AVG INCOMING MATH NCE (END OF 5 TH GRADE)	0.20	0.10	.062
K8	1.76	1.26	.171
YRS OF MATH WSR	1.18	0.57	.043
YRS OF OTHER CSR	0.47	0.29	.117

Figure 1. 7th to 8th Grade Math NCE Gains on SAT9, 1999-2000



¹ Of the 14,448 fifth graders with PSSA math scores in 1997, by 2000 more than 5000 students had left the school system, not been promoted, not enrolled in a regular eighth grade school, or were missing an eighth grade math score for some other reason, and had to be excluded from the analyses. In particular, 2,999 of the original students with PSSA math scores in 1997 were no longer enrolled in Philadelphia public schools 3 years later. The fifth grade math scores of these students did not differ significantly from those of the students who remained. A total of 884 students had not been promoted to grade 8 by 1999-2000 (and, not surprisingly, their fifth grade scores were significantly lower than those of students who reached eighth grade on time). Of the 10,565 students with fifth grade scores who were in eighth grade in Philadelphia in 1999-2000, PSSA math scores were available for 9768 students. The 797 eighth grade students who were missing PSSA math scores in 2000 scored significantly lower than other students in fifth grade. An additional 195 students were excluded from analyses because they were not attending regular schools for middle grades students, but as a group, these students average eighth grade math score did not differ significantly from those of included students. The rest of the students excluded from the final analyses were missing demographic data, but did not differ significantly from the included students included on fifth or eighth grade PSSA math score.