Final Report to the Institute of Education Sciences
October 1, 2008

Evaluation of the Computer and Team Assisted Mathematical Acceleration (CATAMA) Lab for Urban, High-Poverty, High Minority Middle Grade Students (R305F050223)

Robert Balfanz
Allen Ruby
Doug Mac Iver

Center for Social Organization of Schools
Johns Hopkins University

This project entailed a three-year efficacy evaluation of the Computer and Team Assisted Mathematical Acceleration (CATAMA) Lab developed by the Center for Social Organization of Schools at Johns Hopkins University. The CATAMA Lab was proposed as an immediate and practical approach to addressing the different types of math deficits held by students at urban high-poverty schools. The Lab required only 1 teacher per school reducing staff and professional development requirements. It used of multiple instructional techniques (including individualized computer instruction, direct instruction, pair and team learning, and individual instruction) to teach math concepts and skills. By taking the place of an elective it allowed students to continue with their on-grade math class. For a more detailed description of the Lab see Appendix 2.

The original goal of the project was to establish the Lab at three urban schools serving high-poverty high-minority middle grade students (grades 5-8). Students underperforming in mathematics (as established by district standardized tests) were to take a trimester course of study in the Lab to increase their knowledge of math concepts and skills taught by a regular math teacher receiving extensive ongoing professional development. Students were to take the Lab as an elective course while continuing with their regular math class. From each school's pool of students eligible to participate, students were to be randomly assigned to take the Lab. An implementation analysis was to measure the teaching of the concepts and skills to be taught in the Lab. To evaluate the impact of the intervention, students' math achievement, as measured by standardized math tests, was to be compared to eligible students not assigned to the Lab. This report discusses the project in three sections:

1) A comparison of the actual project with the planned project
2) The descriptive results from the project
    a. Description of the sample
    b. Description of implementation of the CATAMA Lab
3) The evaluative results from the project

I. Comparison of the Actual Project with Its Original Design

Originally, the CATAMA Lab was to be established and maintained for two years at three neighborhood middle schools in Philadelphia serving high-minority low-income student populations. All three schools had agreed to take part in the study and to provide: 1) a lab teacher plus time for their professional development, 2) a room, 3) the requisite number of computers, and 4) randomization of eligible students into the Lab or another elective class, and 5) scheduling for the class and for the assessment of treatment and control students. In addition, the Philadelphia school district gave permission for the study. Using the grant funds, CSOS agreed to provide: 1) professional development including a 2 day workshop before the start of the Lab and weekly in class support from an experienced ex-Lab teacher, 2) the math software (Larson's Pre-Algebra), 3) additional lessons and classroom materials (e.g., overhead projector, student boards and markers, posters, etc.). The study was to focus on 5th and 8th grade students because these students were tested each spring by the district using the CTBS Terra Nova math test.

These test scores were to be used in the project as the outcome variable measuring student achievement in math. The schools were to hold 5 CATAMA Lab classes a day for one trimester (the schools were on a trimester grading period). New classes would be held each trimester for a total of 15 classes a year. Each CATAMA Lab class was to have 15 to 18 students. However, the district's standardized testing would occur halfway through the last trimester so only students from the first two trimesters a year would be included in the study. The expected number of Lab students was at a minimum: 15 (students per class) X 10 (classes a year) X 3 (schools) X 2 (years) = 900 treatment students. A similar number of control students were expected as students were to be randomly assigned to the CATAMA Lab or an elective class by flipping a coin.

As seen in the year-by-year discussion below the original plan was overtaken by technical, school, and district decisions and events. The overall impacts were: 1) the actual sample size was half of the expected sample size but it proved large enough to achieve the expected power for the analysis, and 2) the original three schools were unable to stay in the study for the full three years and so an additional 3 schools were recruited into the study. The details are discussed below and Table 1 provides a summary of the schools and dates of Lab implementation.

Year 1

In Year 1, these goals were met but with three modifications: 1) a smaller sample size than expected, 2) inclusion of students from additional middle grades besides 5th and 8th, and 3) faster progress in data collection than expected. Before discussing these exceptions, this report notes that the other planned steps in the evaluation all took place as planned. The three schools that signed up to take part in the study before the grant was received did take part in the study. A team of project and school personnel identified the eligible students at each school based on their previous year's standardized test scores and project personnel randomized them by grade and math class into Treatment and Control groups. Treatment students were scheduled to take the CATAMA Lab while Control students were scheduled into another elective class. Each school chose a Lab teacher and the three Lab teachers received the initial professional development necessary to lead their Labs. The Labs were implemented over the first trimester although two of them started late for technical reasons. First, the company that marketed the software had been sold and the new company was not ready to ship the software until the fall of Year 1 (the 3rd school had the software already in place). Second, at School 2, the original principal and Lab teacher left the school in September. A new Lab teacher had to be found and trained and a new Lab had to be established as the new principal wanted to use the original Lab for computer instruction. Weekly in-class professional development was provided to the Lab teachers. In addition, a Lab Observation protocol was established and observations were made of Lab implementation.

The first modification to the original plan was a smaller sample size than planned. In Year 1, 460 students took part in the study rather than the planned 900 students. This occurred because of several policy decisions and technical problems.

1. Fewer Lab classes were held than planned: At Schools 2 & 3 fewer than the planned 5 Lab classes were held. At School 2, the new Lab teacher was also the school's math coach and only had time for 2-3 Lab classes. At School 3, 4 classes fit the schedule better than 5 classes.

2. Cycle 2 was not successful at two schools and was only successfully completed at School 3. At School 2, it was ended before the half way point by a computer failure that was not rectified in time to complete the cycle. At School 1, the principal decided to keep the Cycle 1 5th & 8th graders in the Lab for test preparation. His decision was made in response to a district decision (announced in the winter) to count only those grades' test scores for AYP calculations.

The second modification was the inclusion of 5th through 8th grade students in the study rather than only 5th and 8th grade students. This modification was made partly in response to requests by principals partly based on a belief that the district would begin to count the standardized test scores of students in all grades for AYP calculations (a decision that was not made for that year) and partly in response to the scheduling constraints of the schools (because of the elective scheduling, it was not possible to schedule all eligible 5th and 8th graders for CATAMA). Because the CATAMA Lab was developed for all the middle grades, this modification seemed to strengthen the validity of the study and because the scheduling difficulties would have reduced the sample available, the project team accepted it.

The third modification concerned the student and test score data collection. Originally, this data was to be obtained from the school district's records: Year 1 data was expected to be availably by the late fall of Year 2. However, in August-September of Year 1, the district began a public discussion of whether it would continue using the CTBS TerraNova math test, which the project was to use as a measure of student math achievement. To avoid the loss of this achievement measure, the project with the schools' agreement pre and post-tested the students in the study using the CTBS TerraNova. Testing was carried out by project personnel with teachers in the classroom to help with classroom management but not with testing issues. Once we began collecting the information needed for pre and post-testing it became obvious that student demographic and attendance data could be collected at the same time. This had four benefits: 1) we obtained the data sooner, 2) the data were very clean, and 3) attendance rates could be calculated for the study period rather than for the whole year as the District data would have been provided, and 4) we had a pre and post test from the same school year and so did not have the problem of summer loss. As a result of having the data sooner, we were able to do a set of initial analyses of Year 1 which showed a significantly positive effect of the Lab on student achievement (with an effect size of .26) that we presented at a poster session at the June 2006 IES Summer Research Conference.

Year 2

In Year 2, further modifications had to be made to the original goals. Because of budget cuts by the School District of Philadelphia, Schools 2 & 3 were no longer able to support a Lab and they dropped out of the study. The study continued at School 1 where the Year 1 Lab teacher had been promoted to a non-teaching position and a new Lab teacher had to be trained.

To replace the two schools that dropped out, the study was also expanded to two new schools: 1) School 4 - a Philadelphia high school that requested the Lab during the first semester for its 9th grade Algebra 1 students who were lacking pre-algebra skills: 1) School 5 - a middle school on an Indian reservation in MN with whom we already had a working relationship. School 4's population for the study was similar to those of Schools 1-3 as it was an urban neighborhood high school serving a student population that was under-prepared to succeed in algebra – the only difference was that the students were a year older than the 8th graders already in the study. That difference was not seen as affecting the theory behind the Lab and the inclusion of the school was seen as a way to test whether the Lab could succeed in a high school environment. School 5's population was different from Schools 1-3, as the school primarily served a Native American population in a rural area. Again, however the underlying theory of the Lab was expected to hold for these students as they were attending middle school and under-prepared for algebra.

The equipment and training costs of switching schools were minimal as we transferred the site licenses for the software to the new schools and the schools paid for their teacher's attendance at the professional development. However, there were higher travel costs for the support of the MN school and as a result technical support was reduced to once a month (versus once a week at the Philadelphia schools).

At all three Year 2 schools, a team of project and school personnel identified the eligible students and project personnel randomized them by grade and math class into Treatment and Control groups. Treatment students were scheduled to take the CATAMA Lab while Control students were scheduled into another elective class. The three Lab teachers received the initial professional development necessary to lead their Labs. In-class professional development was provided on a weekly basis to the Lab teachers at the two Philadelphia schools and on a monthly basis to the school in MN (with the Lab facilitator maintaining contact with the Lab teacher through email and phone contact in between visits). The Lab Observation protocol developed in Year 1 was used to make observations of Lab implementation.

Two cycles of the Lab were held at School 1. One cycle was run at the other two schools. At School 4, the eligible 9th graders were divided into two groups. The first group took the Lab during first semester and was the treatment group. The second group took an elective during first semester and served as the control group then took the Lab during second semester. This design was done at the school's request as they wanted all their under-prepared students to take part in the Lab. School 5 decided to join the study after the beginning of Year 2 and so the Lab started up in its second semester and so only 1 cycle was held. As a result, the project had obtained about one-half the planned sample size. However, at this point we have collected enough middle grade results to reach our minimum power requirement of .80 given an estimated effect size of .20.

We continued the modification made in Year 1 of switching from the school district's provision of test score and demographic data for each student to doing our own pre and post-testing using the CTBS TerraNova Survey and collecting the other data directly from the schools. Just as in Year 1, this approach had three benefits: 1) we obtained the data sooner, 2) the data were very clean, and 3) attendance rates could be calculated for the study period rather than for the whole year as the District data would have been provided, 4) we had a pre and post test from the same school year – no summer loss.

One other modification of the implementation came out of Year 1 and Year 2: keeping students in the Lab for 1 semester rather than 1 trimester. It became clear from the Philadelphia work that the scheduling demands on both the school staff and on students for have three Lab cycles a year were too burdensome and it was simpler to keep students in the Lab for one semester. In the high school and the MN schools this was a natural change as they worked on a semester system.

We continued with the modification of pre and post-testing the students ourselves using the CTBS TerraNova Survey. The benefits of this were obvious for the new schools. For School 4, we tested the 9th graders and the results could be put on the same scale as those of the middle grade students. For School 5, we tested the middle grade students using the same test that the other students in the study were receiving rather than use the MN state test. In Year 2, we have analyzed the high school data where the study ended after the first semester and found an effect size of .63 on the 9th grade Lab students. These results show that the positive benefits of the Lab for middle grade students' math achievement found in Year 1 were transferable to 9th graders taking Algebra 1 (within the much different institutional structure of a high school). The results for the 9th grade were presented in a poster session at the June 2007 IES research conference.

Year 3

In Year 3, due to continuing budget cuts in the district, School 1 decided it could not maintain a Lab and dropped out of the study. School 5 remained in the study and a new school (School 6) with which we had an ongoing relationship was added to the study. School 6 was a middle school in the San Antonio School district serving a predominantly Hispanic student population. The underlying theory of the Lab was expected to hold for these students as they were attending middle school and under-prepared for algebra. Only 1 cycle was run in Year 3 at the schools' request: the eligible students were divided into two groups. The first group took the Lab during first semester and was the treatment group. The second group took an elective during first semester and served as the control group then took the Lab during second semester.

The modifications made in Years 1 & 2 were maintained in Year 3. The addition of School 6 raised travel costs and to offset them, technical assistance was provided on a monthly visit with the Lab facilitator maintaining contact with the Lab teacher through email and phone contact in between visits. The Lab Observation protocol developed in Year 1 was used to make observations of Lab implementation.

Table 1. Summary of Implementation Start and Stop and Testing Dates

Year: Cycle	School: Grades	Initial Lab Teacher Training	Lab Started	Lab Ended	Pre-test	Post-Test	Class Days Between Testing
Year 1: Cycle 1	1 6th & 8th	Summer 2005	9/6/05	1/26/06	9/19/05	1/26/06	88
	2 6th & 8th	Summer 2005	11/1/05	2/17/06	9/28/05	2/17/06	53
	3 7th & 8th	Summer 2005	10/17/05	1/13/06	9/20/05	1/11/06	54
Year 1: Cycle 2	1		na	na	na	na	
	2 5th & 6th	**	3/6/06	***	2/16/06	not held	
	3 5th & 6th	**	1/17/06	4/6/06	1/10/06	4/5/06	50
Year 2: Cycle 1	1 6th-8th	Summer 2006	9/11/06	12/22/06	9/12/06	1/12-1/15/06	63
	4 9th	Summer 2006	9/11/07	End of year	9/19/06	1/23 -1/26/07	63
Year 2: Cycle 2	1 6th-8th	Summer 2006	1/2/07	5/18/07	1/3-1/4/07	5/14-5/18/07	82
Year 2: Cycle 2	5 5th-8th	January 2008	1/23/08	5/08/07	1/15-1/18/07	5/7-5/10/07	55
Year 3: Cycle 1	5 5th-8th	Summer 2008	9/5/07	1/17/08	9/11-9/14/07	1/16/08	76
	6 7th & 8th	Summer 2008	7th: 9/5/07 8th: 9/10/07	1/29/08	9/4-9/7/07	1/29/08	58

* Subtracts holidays, professional development days, snow days, and test days.
** Initial training held before Cycle 1
*** Lab ended when school lost software.

II. Descriptive Data

A. Description of the sample

During the study, 1090 students were found eligible to take part. Of these 985 students completed the study (took the pre and post-test) and 105 students attrited from the study.

Students in the Study

Table 2 breakdowns the sample of 985 students by school, cohort, grade, gender, and race/ethnicity. School 1 contributed the most students because the Lab ran for 3 cycles there (versus two cycles at Schools 3 & 5, and one cycle at Schools 2, 4 and 6). The majority of the students came from Cohorts 1 & 5 and from 8th and 6th grades. There were more girls than boys in the study. Blacks and Hispanics made up the majority of the students.

Table 2. Composition of Sample

Breakdown sample by:	Number of Students
Total	985
School	
School 1	375
School 2	55
School 3	168
School 4	62
School 5	137
School 6	188
Cohort*	
Cohort 1	352
Cohort 2	84
Cohort 3	159
Cohort 4	135
Cohort 5	255
Grade	
5th	66
6th	225
7th	167
8th	465
9th	62

Gender	
Female	541
Male	443
Race/Ethnicity	
Asian	60
Black	300
Hispanic	432
White	48
Other/American Indian	143

*Cohort represents semester of implementation during the 3 year study period.

There were 552 treatment students and 433 control students. Table 3 compares the treatment and control groups on their composition by subgroup. In the school comparisons, we see that School 3 had a statistically significant larger percentage of treatment students than control students but none of the other schools had such a difference. For the cohort comparisons, we see that Cohort 2 had a statistically significant larger percentage of treatment students while Cohort 5 had a significantly larger percentage of control students. By grades, the control group had a significantly larger percent of 5th and 8th graders. The treatment group had a significantly larger percent of Asians and Blacks and a significantly smaller percent of Hispanics. The treatment group had a significantly larger percentage of females.

Table 3. Subgroup Comparison of Treatment and Control Groups

	Treatment N = 552	**Control N = 433**	**Difference**	**P-Value**
School 1	38%	38%	0	.998
School 2	5%	6%	-1	.616
School 3	21%	11%	+10	.000***
School 4	7%	6%	+1	.735
School 5	13%	16%	-3	.157
School 6	16%	23%	-7	.007%
Cohort 1	37%	35%	+2	.526
Cohort 2	11%	6%	+5	.005**
Cohort 3	17%	15%	+2	.301
Cohort 4	14%	14%	0	.949
Cohort 5	22%	31%	-9	.001***
Grade 5	8%	6%	+2	.191
Grade 6	24%	22%	+2	.453

Grade 7	18%	16%	+2	.355
Grade 8	44%	51%	-7	.033*
Grade 9	7%	6%	+1	.740
Asian	8%	4%	+4	.009**
Black	33%	27%	+6	.037*
Hispanic	40%	49%	-9	.009**
White	6%	4%	+2	.119
Other	13%	17%	-4	.110
Female	59%	50%	+9	.008**

A comparison of pre-test scores shows no significant difference between the treatment and the control group in their math skills. Table 4 provides these results using several different measures derived from the raw scores: scaled scores, national percentiles, grade equivalents, and normal curve equivalents. The national percentile scores show that on average the eligible students were performing at the 33rd to 34th percentile compared to the average U.S. student.

Table 4. T-tests of Mean Pre-test scores

Pre-test Score	Treatment n = 552	Control n = 433	Difference	P-Value
Raw Scores	14.6	14.8	-0.2	.512
Scaled Scores	642	646	-4	.248
National Percentiles	33.1	34.6	-1.5	.278
Grade Equivalents	6.0	6.2	-0.2	.147
Normal Curve Equivalents	38.1	39.1	-1.0	.359

These descriptive data show that the treatment and control students started with the same average level of math achievement. There were several differences between them regarding the percentage of each from certain schools, cohorts and races/ethnicities but these provide neither group with an apparent advantage especially when the level of prior achievement is similar.

<u>Attrited Students</u>

One hundred and five students dropped out of the study from the original 1090 students. This represents 9.6% of the study's original sample which is a low percentage given the high rate of mobility found in these types of schools. Table 5 compares the attrited students from those in the study. The table shows that a statistically significant larger percentage of attrited students came from School 4 and from 9th grade (these are equivalent since only School 4 contributed 9th graders to the study) and a lower percent from School 3 and 5th grade than the percentage from students in the study. In addition no Asian students attrited. However, there was no difference in the mean pretest score between the attrited students and students in the sample.

Table 5. T-tests comparing students who withdrew from sample to those in study

	Attrition N = 105	Sample N = 985	Difference	P-Value
School 1	30%	38%	-8%	.095
School 2	8%	6%	+2%	.465
School 3	4%	17%	-13%	.000***
School 4	16%	6%	+10%	.009*
School 5	16%	14%	+2%	.550
School 6	26%	19%	+7%	.100
Raw Pretest Scores	14.3	14.7	-0.4	.408
Scaled Scores	646	644	+2	.684
National Percentiles	34.2	33.8	+0.4	.834
Grade Equivalents	6.3	6.1	+0.2	.432
Normal Curve Equivalents	38.2	38.5	-0.3	.862
Grade 5	2%	7%	-5%	.002**
Grade 6	19%	23%	-4%	.330
Grade 7	15%	17%	-2%	.630
Grade 8	48%	47%	+1%	.868
Grade 9	16%	6%	+10%	.009**
Asian	0%	6%	-6%	.000***
Black	29%	30%	-1%	.685
Hispanic	52%	44%	+8%	.101
White	6%	5%	+1%	.708
Other	13%	15%	-2%	.740
Female	46%	55%	-9%	.088
Treatment group	64%	56%	+8%	.101

The attrited group include 67 treatment students and 38 control students (12% and 9%). To determine if there was a difference in the students who attrited from the treatment group versus the control group, the mean of these students were compared. Table 6 shows no difference in the mean pretest score of treatment attrited students and control attrited students (as measured in NCEs). That the treatment and control attrited students did not significantly vary by prior achievement and that the treatment and control students in the sample also did vary by prior achievement gives us confidence that differential attrition was not the cause of the study's results (that a greater percentage of students resistant to improving their achievement left the treatment group thereby causing any greater gains for treatment students).

Table 6. T-tests of prior achievement for those students who withdrew from study

Pre-test score	**Treatment n = 67**	**Control n = 38**	**Difference**	**P-Value**
Normal Curve Equivalents	38.3	38.0	+0.3	.942

B. Description of implementation of the CATAMA Lab

To track implementation, we used a monthly observational checklist implemented by a single Hopkins employee familiar with the Lab and how it should be run. The checklist addressed the: 1) availability of all necessary materials, 2) use of the teaching routine, 3) level of differentiated instruction, 4) promotion of teamwork, 5) use of motivational practices, and 6) level of student engagement.

Implementation at each school was high in that students attended the Lab for the expected period of time. The use of computerized instruction was also high, although there were interruptions due to hardware problems, with students successfully working in teams at different paces to fill in gaps in their math knowledge. Student engagement appeared high and quickly adapted to the routine of instruction reducing time spent by teachers on classroom management.

However, the teacher instructional components (whole class and small group instruction, and motivational practices) were not as well implemented. In part, this was due to the teachers having to learn a different instructional approach from their normal instructional methods that were often based on lecture, reading from the book, doing problems, or asking students to do seatwork in class. The original plans for the project assumed that the Labs would continue in the original three schools for all three years and that the Lab teachers would, on the whole, also remain through the same through the three years. Under this assumption, Lab teachers would become more experienced over the years and their Lab instruction would improve. But only in School 5 did the Lab teacher teach the Lab for two years and she did become more skilled in the second year. In School 1, the first left teacher left for another position in the second year, and Schools 2-4 and 6 were in the study for only one year. As a result, we were not able to determine whether teacher instruction improved over time.

We also found two other contributing factors to weaker than expected instruction: 1) overextension of teachers, and 2) teacher self-discipline. Because instruction in the Lab is class and student specific, it requires ongoing preparation by teachers, especially ones new to the Lab. When teachers took on additional educational duties, by their own volition or by school assignment, they lost their preparation time. Table 7 shows how almost every teacher was engaged in some other educational pursuit. The additional assignments placed on them by their school (e.g., act as a math coach for the entire math faculty or take on a new algebra 1 course) occurred during the school year forcing the teacher to do their planning while they taught rather than preparing before school began. In response to the additional workload (be it from the school or to further their career) many of the Lab teachers reduced their teaching and motivational activities in favor of more computer instruction allowing them to work on their other job-related requirements or at times to even relax.

These factors impeding implementation would be typical for any educational program implemented by school personnel and supported by an outside organization especially during the first year of a study in schools (and a district) serving high-poverty high-minority populations. Although, it might be expected that if a district adopted the Lab, there would be greater stability of the Lab and its instructor within each school allowing a more realistic view of whether instruction improved over time (and with it student achievement). While the use of all the instructional components were not as high as desired, their low level of use, the expected level of use of the computerized instruction, plus the regular holding of Lab classes for all the treatment students as scheduled combine to give a level of implementation acceptable for studying the impact of the Lab on student achievement.

Table 7: Quality of Teacher Implementation

Year	School (Cycle)	Other Teacher Duties
Year 1	1 (Cycle 1)	Engaged in training to become an Assistant Principal
	2 (Cycle 1)	Math coach for entire school
	3 (Cycle 1 & 2)	Teaching an algebra 1 course (never taught algebra before)
Year 2	1 (Cycle 1 & 2)	Started graduate school for an education degree.
	4 (Cycle 1)	
	5 (Cycle 2)	Completing required course of study for advanced credential.
Year 3	5 (Cycle 1)	
	6 (Cycle 1	Non-certified teacher taking courses for certification.

III. Impacts of CATAMA Lab

As noted in Section I, earlier analyses found that the CATAMA Lab had a positive impact on student gains in pre-algebra math achievement (as measured by the CTBS TerraNova) for the students in the first year of the study and for 9th grade students who took part in Year 2 of the study. The discussion of first year results can be found in Appendix 2 (a paper submitted to the American Journal of Education) and a discussion of the 9th grade results can be found in Appendix 3 (a paper presented at AERA 2008).

Here we discuss the results from all three years of the study. Table 8 presents the findings from the comparison of means for the treatment and control. As the study was an experiment with randomization at the student level, a simple t-test of the means provides us with an estimate of the impact of the Lab. The table shows that Lab students made a statistically significant gain in pre-algebra math achievement from their semester attendance in the Lab. For example, while both treatment and control students rose in the national percentile (i.e., their achievement increased relative to the national performance on this assessment) Lab students rose by more than twice as many percentiles (10 versus 4).

Table 8. T-tests of Post-test Sores and Gains in Post-test Scores

	Treatment N = 552	**Control N = 433**	**Difference**	**P-Value**
Post-test: Raw Score	17.5	16.5	+1.0	.002**
Post-test: Gain in Raw Score	+2.9	+1.6	+1.3	.000***
Post-test: Scale Score	664	657	+7	.014*
Post-test: Gain in Scale Score	+21	+12	+9	.000***
Post-test: National Percentile	43.4	39.0	+4.4	.003**
Post-test: Gain in National Percentile	+10.3	+4.4	+5.9	.000***
Post-test: Grade Equivalent	7.2	6.8	+0.4	.012*
Post-test: Gain in Grade Equivalent	+1.2	+0.6	+0.6	.000***
Post-test: NCE	45.3	42.2	+3.1	.003**
Post-test: Gain in NCE	+7.2	+3.1	+4.1	.000***

In addition to the bivariate analysis, we applied a model-based approach to address possible differences that could occur after assignment. For example, we would expect attendance to affect student achievement and attendance occurred after the random assignment. In addition, because the randomization was not made within blocks of individual characteristics, a model controlling for these characteristics can more accurately estimate the treatment effect.

We used an OLS regression model with change scores as the dependent variable to model the effect of the Lab. With only 1 lab teacher per school and 6 schools, we did not have enough cases for a hierarchical model. However, the inclusion of the dummy variables representing the schools and teachers in our model controlled for all unobserved characteristics of the schools and teachers more appropriately than a hierarchical model because it allowed for the correlation between these dummy variables and the other regressors, including the treatment. The dependent variable is y_i, the gain in test score for student i. The key independent variable, denoted T_i , takes the value of 1 for the Lab treatment students and O for the control students,. Other control variables measure the characteristics of the students, their regular math teacher and their Lab teacher. Student characteristics (represented by X_i) include: grade level, gender, and race/ethnicity (Asian, Black, Hispanic, White, Other), and attendance rate. Differences among the schools were controlled for using a dummy variable for each school (S_i) which also effectively controlled for Lab teacher differences because there was only one Lab teacher per school. The model is expressed as:

$$y_i = \beta_0 + \beta_1 T_i + \beta_3 X_i + \beta_5 S_i + \varepsilon_i \quad (1)$$

All the independent variables, except the treatment, were centered around their specific mean to provide a clearer interpretation. The coefficients for the independent variables remain the same with or without centering. After centering, the intercept, $\beta_{0,}$ captures the average gain for the control group and can be interpreted as the gain in test score for the typical student at the mean of each covariate who did not attend the Lab. β_1 captures the average Lab effect in terms of an additional gain for the Lab group when controlling for the other covariates and can be interpreted as the additional gain in test score due to the Lab for the average student. If β_1 is significantly positive and substantial, we have evidence that the Lab successfully increases the Lab students' math achievement as compared to the control group.

Table 9 shows the results from the estimation of the model using NCEs as the measure of student performance for both the pre and post-test. A positive significant coefficient of 3.5 was found for Lab attendance. This was smaller than the gain found for the bivariate comparison because we have partialled out any contributions of factors not controlled for by our original randomization of students.

This impact of the Lab can be converted into an effect size of .10. Using a composite of math standardized tests, Bloom, Hill, Black and Lipsey (2006) found that one year of regular math instruction had an effect size between .19 - .41 for middle grade students (declining as grade increased). In other words, students spending an additional 15-20% of time in math instruction in the Lab make achievement gains equivalent to spending about 25% to 50% a year in their regular math class.

Other significant variables included: 1) students at School 2 did worse than those at School 1 while students in Schools 5 & 6 did better than those at School 1 (used as the comparison school in this model), 2) students in the 6th and 7^{th} grades had greater gains than those in 8^{th} grade (used as the comparison grade), 3) and students with greater attendance had greater gains. The R^2 was .38 which is relatively high for change models. The covariates of days in CATAMA, regular math teacher credentials, and lab teacher credentials were not found significant and were dropped from the final model

Table 9. Estimates of Lab Impact from OLS Regression

Variable	**Coefficient**	**Effect Size**	**P-Value**
Constant	2.917		.533
Prior NCE	.523	.530	.000***
School 2	-4.408	-.062	.024*
School 3	.957	.022	.500
School 5	6.875	.145	.016*
School 6	3.052	.071	.026*
5^{th} grade	2.553	.039	.179
6^{th} grade	4.137	.106	.000***
7^{th} grade	3.523	.080	.004**
9^{th} grade	3.635	.054	.061
Asian	.521	.008	.788
Black	-.250	-.010	.782
White	4.218	.056	.053
Other	-2.004	-.043	.478
Female	.262	.008	.757
Attendance	.169	.095	.000***
CATAMA Lab (Treatment)	3.453	.104	.000***

Note: School 4 is not included as all its students and only its students are 9^{th} graders so it is identified by the 9^{th} grade covariate

- Baseline case for a Male, Hispanic, 8^{th} grader from School 1
- R^2 = .376; F-statistic = 37.483; P-Value = 0.000***

IV. Concluding Comments

The results from the CATAMA Lab evaluation provided both supportive evidence for the program as an intervention but also identified a need to further address some of the teaching instruction issues surrounding its implementation. A significant positive impact was found on Lab students' gain on a pre-algebra test compared to control students who did not participate in the Lab. The gain appeared larger than would be expected from spending the additional time in students' regular math class (this comparison was made by the effect size of the gain compared to the effect size of spending a year in school and not by comparing control students who had spent the time in regular math class – an option not available in the study schools). As an aside, this type of comparison has been a sticking point in getting the results of the study published. The first year results (see Appendix 2) were submitted to two journals. In both cases, the reviewers split with some in favor and others saying the results could be due to extra time on math rather than the Lab itself. A third revision is being prepared with the argument that extra time doesn't always result in greater gains.

The results of the fidelity of implementation portion of the study are somewhat sobering. Computer instruction can misused by teachers who want to spent Lab time on non-Lab activities. Rather than use the time when students were working on the computer to work with individuals or small groups or to observe students to identify where they are having difficulties, some teachers spent too much of this time working on other education-related projects (either their own or those assigned by schools) or relaxing/socializing.

Addressing this issue could take two forms. First, perhaps the evidence of a Lab impact could be used to convince a single district to commit to maintaining the Lab as a longer-term (perhaps 3 year) intervention using the same set of teachers. With time, the teachers would better understand the importance of the Lab routine, become more skilled at it, and achieve greater fidelity of implementation, theoretically leading to greater student gains.

Second, perhaps more forceful teacher professional development could be applied stressing the role of the teacher throughout the Lab. This training could be accompanied by providing a more scripted approach for first-year Lab teachers that included a pacing schedule for each class. Such an approach conflicts in part with the Lab's focus on the teacher identifying and addressing student needs as they arise. But perhaps this skill is better learned after the teacher has experience in the basic methods of teaching the Lab and would be better stressed in the second year of Lab implementation after the teacher has a better understanding of his or her roles for each of the instructional methods used in the Lab.

Appendix 1: Implementation Checklist

CATAMA Lab

School:

Grades Serviced: **Cycle:** **Date**:

Daily warm-up: (check) ○ **Mental Math** ○ **Problem of the day** ○ **Journal writing** ○ **Vocabulary building** ○ **game**	**Explanation of daily warm-up:**
Computer Software: (check) ○ **Larson's Prealgebra** ○ **Skillsbank 4** ○ **Cornerstone** ○ **Destination math** ○ **Games**	**List module or subtopic students:**
Room Setup and Mat'l : (check) ○ **Computer tables with 10 computers (PC or laptop)** ○ **Centers** ○ **Teams** ○ **Partnerships** ○ **Number line** ○ **Place-value chart** ○ **Word Wall** ○ **Student Folders** ○ **Wipe-off boards** ○ **Calculators** ○ **Resource text** ○ **Games** ○ **Fact Cards** ○ **Overhead Projector** ○ **Chalk Board or other**	**Additional material:**
Routine: ○ **Warm-up:** **5-minute math** **Problem of the Day** **Math game** **Vocabulary review** **Journal writing** ○ **Computer Assignment** **Pre-test**	

Module assignment **Posttest** ○ **Whole-group Instruction** ○ **Small-group Instruction** ○ **Mini-lessons** ○ **Centers: games, Open-ended questions, Problem-solving strategies, journal-writing, fact building, basic skills fluency, extra practice, vocabulary building** ○ **Vocabulary review** ○ **Closure: review vocabulary of module, fact review, math concept review** ○ **Clean-up and dismissal**	
Differentiated Instruction: (check) ○ **Manipulatives** ○ **Basic fact practice/tests** ○ **Disecting word problems** ○ **Operational Vocabulary** ○ **Small-group instruction** ○ **Algorithm procedures**	**Explain:**
Student Engagement: (Check) ○ **100%** ○ **75%** ○ **50%** ○ **less than 50%**	**Reason for the indicated percentage of student engagement:**
Teamwork: ○ **Students help each other to solve problems and determine strategies to solve problems** ○ **Students give answers instead of discussing strategies** ○ **Students ask for help from other teams and partnerships** ○ **Students only ask for help from partner** ○ **Teacher provides incentive for team effort.**	**List observed teamwork activities or incentives encouraging teamwork.**
(check) ○ **Teacher encourages students to use other resources for**	

clarification of concept ○ Students are encourage to refer back to instructional page when necessary. ○ Students are encouraged to use visuals in the room ○ Teacher provides mini-lesson when needed ○ Teacher encourages students to ○ Teacher encourages both students in partnership to solve each problem, discuss answers then key in answer choice ○ Teacher provides incentives and motivation: - Certificates of mastery - prizes - games List any observed: - parent notes - praise - other	

Appendix 2: Year 1 Paper

Improving Math Achievement of High-Poverty Urban Middle Grades Students: An Extra-Help Math Lab Approach

Allen Ruby and Robert Balfanz

Abstract

During the middle grades, students at urban schools serving high-poverty high-minority populations often fall severely behind in math achievement. While benefiting from current reform efforts to improve instruction, these students also require extra help to close their math skill and knowledge gaps. We report the results from a randomized experiment for an extra-help math lab that uses a combination of teacher, peer and computer instruction, to address the particular gaps of each student. Lab student gains were double those of non-Lab students with the gains similar to those obtained from a year of regular math class. The results provide evidence for the importance of extra-help programs that address individual student needs while being practical for schools.

Improving Math Achievement of High-Poverty Urban Middle Grades Students:

An Extra-Help Math Lab Approach

For many high poverty students, the middle grades are where achievement gaps in mathematics become achievement chasms. Nearly all high poverty students enter kindergarten with the most basic mathematical knowledge at hand--they can count and recognize basic shapes (West, Denton, & Reaney, 2000), but many end middle school ill-prepared to succeed in a rigorous sequence of college preparatory mathematics courses in high school (Author, 2002).

National and international comparisons of student achievement indicate that it is between 4th and 8th grade where U.S. students in general, and minority and high poverty students in particular, fall rapidly behind desired levels of achievement (Beaton et al. 1996; Schmidt et al. 1999). In nearly all of the nation's states there is a 30 to 50 percentage point difference between white students and the largest minority group in the percent of students scoring at basic on the 8th grade NAEP exam (Blank & Langesen, 1999).

Nationally these differences have recently been replicated for minority versus white students and low-SES students versus higher-SES ones by the Program for International Student Assessment (National Center for Educational Statistics, 2004). Many of these minority students, in turn, are concentrated in high poverty urban schools. For the students attending these schools, and the nation as a whole, low mathematical proficiency at the end of the eighth grade has serious consequences. The ability to succeed in college preparatory mathematics courses in high school has been linked to success in post-secondary schooling and to life-long opportunities for success (Pelavin &

Kane, 1990; U.S. Department of Education, 1997). In addition, large concentrations of poor and minority students who receive weak academic preparations in their middle school years help to create neighborhood high schools in our nation's largest cities that function as little more than dropout factories rather than stepping stones to a strong education and upward mobility (Author, 2001).

Many explanations have been offered to explain the middle grades mathematics achievement gap. Weak and unfocused curriculums (Schmidt et al., 1999), shortages of skilled, trained, and knowledgeable mathematics teachers (National Commission on Mathematics and Science Teaching, 2000), unequal opportunities to learn challenging mathematics (Raudenbush, Fotiu, & Cheong, 1998), under-motivated students (Bishop & Mane, 2001), and the turbulence of early adolescence have all been advanced based on credible, if not always comprehensive or incontrovertible, evidence as plausible causes. Each has also brought its own set of reforms. The last decade has seen the advent of more challenging learning standards and higher stakes accountability systems for schools and students, the movement towards smaller learning communities in large middle schools or the conversion of middle schools into K-8's (in efforts to create more personalized learning environments), the spread of research-based mathematics curriculums, and attempts to develop and maintain a stronger corps of middle grades mathematics teachers (Burrill, 1998). Yet, while there has been an overall upward trend in elementary and to some extent middle school mathematics achievement during this period and some notable success in high poverty schools (Chubb & Loveless, 2002), there has been no dramatic and widespread shrinking of the middle grade mathematics achievement gap between more and less advantaged students (Lee, 2002). Even with the most recently reported gains in 8th grade student test scores, including minorities, the gap between schools serving small versus large percentages of economically disadvantaged students remains large (Mullis, Martin, Gonzalez, & Chrostowski, 2004).

Thus existing evidence indicates that in high poverty schools with large concentrations of students with low mathematical proficiencies, higher standards, more accountability, stronger and more focused curriculums, better teachers, and improved teaching and learning environments will all fundamentally be a part of successful efforts to reduce the middle grades achievement gap and prepare more students for success in high school math courses. At the same time, existing evidence also suggests that these efforts may not be sufficient. In high poverty, primarily minority school districts like Philadelphia--the site of this study--where the majority of students enter middle school behind grade level on standardized measure of mathematics achievement and below basic on state assessments, most middle grades students need effective extra help in addition to excellent regular classroom instruction in mathematics in order to close their skill and knowledge gaps and make the transition from elementary mathematics to more complex forms of mathematical thought and practice (Author, 2002).

An illustrative example of this need can be seen in a study of two Philadelphia schools serving high-poverty high-minority populations (Author, 2004). Over the past six years both implemented many recommended practices for improving mathematics achievement in the middle grades including adopting research-based instructional programs, sustained and intensive professional development and teacher support, improved teaching and learning environments, and a high degree of instructional program coherence. Student achievement has increased (Author, 2003). Double the number of students (compared to the district average for similar schools) during the past three cohorts have closed their mathematical achievement gap and leave the 8th grade on or near grade level (Author, 2006). Despite this substantial improvement, half the students in these schools (as compared to three-fourths in the typical high poverty middle school in the district) still leave middle school further behind in mathematics achievement than when they entered.

Consequently, there is a great need to develop and evaluate extra-help programs that can provide critical assistance in the effort to close achievement gaps in the middle grades and prepare students to succeed in standards-based high school math courses. To accomplish this, extra-help programs need to be closely coupled and aligned with challenging standards-based instruction in the regular classroom (Newmann, Smith, Allensworth, & Bryk, 2001), they need to be able to provide substantial assistance to large numbers of students (Author, 1998), and they need to provide a range of mathematical instruction. Existing research on the development of mathematical knowledge and skills during the middle grades indicates that different students will have different extra help needs. Some will need help with the most basic of skills (e.g., multiplication and division), a much larger percent will need help with the intermediate skills and knowledge (such as rational numbers, integers, ratio and proportion,) fundamental to success in pre-algebra and algebra, and still others will need support making the transition to more conceptually complex and symbolically based forms of mathematics (Kilpartick, Swafford, & Findell, 2001).

This study will evaluate an immediate and practical approach to addressing the different types of math deficits held by students at urban high-poverty schools. The

CATAMA Lab incorporates effective multiple instructional techniques to teach math concepts and skills using only one teacher per school, thereby requiring less professional development and no interruption in the existing math instruction, and it can be started up almost immediately in a school while reaching a large percentage of the population in need of assistance. While the Lab is not expected to have the same impact as improving the instruction of all math teachers (nor does it have the corresponding financial and time costs required to do so) it is a means to quickly implement the instruction known to improve students' math knowledge and skills and thereby better prepare underperforming middle grades students for their studies.

The CATAMA Lab

The Computer and Team Assisted Mathematical Acceleration (CATAMA) Laboratory is an elective course for students needing additional assistance in math while they continue in their regular math class. Its purpose is twofold. First, the Lab helps students fill in gaps in math skills and knowledge that they are incorrectly presumed to have already learned in earlier grades. The actual gaps in skills vary widely among students making it very difficult for the regular math teacher to address them. An elective Lab can more efficiently fill the gaps helping students keep up with their grade-level math instruction. Second, the Lab can be used to preview upcoming material from the regular math class. Not only do previews increase the opportunities for low-proficiency students to learn on-grade material but they also help students follow what is being taught in their regular math class reducing the chance that they become lost and give up.

Organizationally the Lab differs from the regular math class. Class size is reduced to 18-20 students selected for their low math standardized test scores. Students attend the

Lab for one to two grading periods (13 - 18 weeks) in place of an elective course (such as art or music). Each section of the lab is dedicated to a particular grade/need combination to facilitate instructional focus and integration with regular math class instruction. For example, the first period class might contain 8th graders struggling in algebra, second period might address 8th graders with weak basic skills (e.g. multiplying positive and negative numbers), while third period could include 5th graders learning to move between decimals, fractions and percentages. The course content then differs by student need and grade level requirements. By combining instruction in math concepts as well as skills the Lab also avoids the traditional criticism leveled at remediation programs of failing to challenge and motivate students because of repetitive practice of low-level skills (Knapp 1995).

The scheduling of the CATAMA Lab as an elective course is key to its success. MacIver (1991) found that the existing evidence suggested that approaches in which struggling students received a substantial extra dose of instruction (e.g. an elective replacement class) were much more effective than less intensive approaches such as before and after school coaching classes. As an elective, the Lab avoids the problems associated with pull-out remediation programs including the inability to keep up with the regular math class, potential differences in teaching between the two classes, and the stigma of being pulled out (Allington, 1991; Bean, Cooley, Eichelberger, Lazar, & Zigmond, 1991). In addition, it also avoids the difficulties of providing specific, systemic skill instruction for struggling students in their regular math classes. As an elective, the Lab can be scheduled throughout the school day. In this way, it can serve large numbers of students and students from all grades with the additional staff requirement of only one teacher. As the course content is not fixed but responds to student requirements, the same student can take the Lab multiple times if needed during the middle grades.

Instructionally, the Lab combines approaches grounded in the theoretical and empirical literature. Each class is taught using three main instructional components: 1) whole class instruction, 2) individual and peer-assisted computer instruction and practice, and 3) individual and small group tutoring. Class begins with the teacher providing approximately 15 minutes of whole group instruction that introduces a skill or concept taught in an earlier grade that students have not yet grasped or previews ones to be introduced in their regular classrooms in the near future. This introduction provides a strong scaffolding for students as it clearly sets out what is to be learned and how it will be learned.

Class continues with 20-30 minutes of individualized and peer-assisted computer instruction building on the individualized extra-help capabilities of computer-based instruction (Macnab & Fitzsimmons, 1999; Abidin & Hartley, 1998). Each Lab has 10 to 15 networked computers loaded with instructional software tailored to their grade and needs. Because different students learn in different ways and have different skill gaps and/or conceptual difficulties, Lab teachers are provided with several computer-based instructional programs, some of which are more skills based and some of which have a more prominent conceptual focus. All of the programs share common features. They provide pre-assessments that tailor the instruction to students' needs, worked/illustrated examples, structured and tiered problem sets, instant feedback, and quizzes and tests that students need to pass at pre-determined levels before the next level of instruction begins. In this study, the Labs relied primarily on the use of Larson's pre-algebra software.

Students of similar skill levels are paired and then teamed with another similar pair in order to take advantage of the motivating and cognitive aspects of peer-assisted learning (Fuchs, Fuchs, Mathes, & Simmons, 1997). Peer- assisted learning techniques are taught so that the student pairs and teams work together. For example, students are taught to “Ask three, before me” or, in other words, first ask their partner, and then their teammates if they don’t understand something before they need to ask the teacher. At times, partners take turns being the ‘reader’ who reads the problem and the “recorder” who inputs the solution. This is done to encourage students to take time to read problems and consider solutions, rather than just attempt to apply the operation they think the problem is calling for (Kilpatrick et al., 2001). The computers are arranged in the classroom so that students sit next to their partners and near their teammates. The teacher also uses motivation activities to help students focus on their work such as making other resources available that students can use to understand a concept, and providing certificates of completion and sending positive notes to parents when students complete a unit. While instruction is peer-assisted, assessment is done individually. Students must pass assessments on their own before moving on to the next instructional level. This motivates partners to help one another so that each will pass and together they can move on.

The time dedicated to computer instruction also provides the Lab teacher the opportunity to provide individual and small group tutoring. There are few effective substitutes for one-on-one or small-group tutoring for students with very large skill deficiencies or knowledge gaps (Wasik & Slavin, 1990). While the class is working on the computers, the Lab teacher can instruct one or several students in a topic they are having difficulty understanding. Tutoring can be formally arranged when the teacher knows a student does not understand a topic or it can take place informally while the teacher circulates during computer instruction and observes a team failing to grasp a skill or concept.

The Lab is taught by an experienced math teacher, viewed by his or her peers as an effective teacher, and familiar with the regular math curriculum at the school giving it the instructional power and flexibility of a strong mathematics teacher (Ma, 1999). In addition, the Lab teacher receives intensive training and ongoing classroom support in the running of the Lab. Before leading a Lab, the teacher receives an initial day of professional development provided by a University Lab facilitator who has experience in both teaching the Lab and supporting Lab teachers. While the training has a theoretical component covering the philosophy and goals of the Lab, the majority of it is focused on practical implementation – use of software, identifying needs of individual students, pairs and the class, and the multiple methods of instruction. Nuts and bolts issues are covered including the Lab materials, lesson planning, setting up and using the computer software, and daily scheduling.

Once the school year begins, the Lab facilitator visits the Lab teacher one day a week to support the teacher and improve their skills. The approach taken by the facilitator differs by the experience of the teacher. The first visits to a new Lab teacher focus on setting up the lab, helping the lab teacher become fully familiar with the software, adapting the work for each class' and each students' needs, correctly assigning the student pairs and ensuring they are working well together on appropriate concepts and skills, and using multiple instructional approaches to help them learn and keep them engaged. The facilitator may lead any of the three components of the class in order to model it for the teacher, co-teach it to give the teacher practice, or observe the teacher and give confidential feedback afterward both on her teaching and the overall running of the Lab. The facilitator is, then, a coach not an evaluator. As the teacher becomes comfortable leading the Lab, the facilitator shifts to giving support through overviews of the Lab's functioning, evaluating classroom needs and small group instruction. The facilitator may work with students to provide the teacher with feedback on which concepts or skills students need additional instruction or practice. Planning, feedback, discussion, and enrichment take place during the teacher's preparation period and lunch so as not to interrupt the Lab. After a year of such support, the teacher is capable of running the Lab on her own however the facilitator's support the next year because it helps the teacher introduce new activities and continue to improve her teaching.

This approach to training the CATAMA teachers is based on the literature's findings that successful professional development is 1) both intensive and long-term on a continual basis, 2) content focused with follow-up training occurring in the context of practice (teaching) through such techniques as monitoring and coaching while also allowing time for reflection and dialogue, and 3) participation should be voluntary and collaboration between researchers and teachers encouraged (Loucks-Horsley, Hewson, Love, & Stiles, 1998; Desimone, Porter, Garet, Yoon, & Birman, 2002).

In sum, the Lab combines the provision of different levels of content with multiple instructional techniques to address individual student math needs, an elective structure that allows it to serve large numbers of students without the drawbacks of other extra-help approaches, a formal structure that can be scaled up in multiple schools but with a flexibility to adapt to changing school and student needs, and an intensive training component that is limited in cost and time requirements with its focus on the Lab teacher rather than the entire math faculty.

Research Questions and Hypotheses

We are interested in determining whether the CATAMA Lab improves student math achievement for students underperforming in math. In this study we examine growth in math achievement during the same year the Lab is taken. Students enrolled in the Lab learn math concepts and skills and apply these in their regular math class. As a result, they should be better prepared for that year's standardized math tests. We expect that students taking the Lab will show greater growth in math achievement than those who do not take the Lab.

We are also interested in whether the Lab has a differential impact on students with different initial levels of underperformance in math. There can be wide differences in this initial level among students taking part in the Lab and alternate hypotheses can be posed whether higher or lower level students might benefit more. From a knowledge and skill point of view, moderately underperforming students need to learn only a few concepts or skills to boost their math achievement and so may benefit more from the Lab than severely underperforming students. From a motivation perspective, severely underperforming students might have given up trying to understand math but when provided with the knowledge and skills to do so their motivation to learn may increase leading to greater achievement gains.

Design

The study uses a pre-test post-test experimental design with random assignment of middle grades students. Three schools in Philadelphia with high-poverty (over 70% school lunch eligibility) and high-minority (85-99% black and Hispanic) student populations volunteered to take part in the study because of their interest in raising their students' math achievement. As the schools were not randomly selected, we cannot claim our results will apply to all high-poverty high-minority schools. At most, we can argue that the results are valid for such schools willing to establish and support a CATAMA Lab.

Student eligibility to take part in the Lab is based on their previous year's math scores. Students who scored between the 25th and 65th national percentiles on the District-given CTBS TerraNova math test were included in the study. Students scoring below the 25th percentile were not included because we have found they often need more individual tutoring to succeed. Students scoring above the 50th percentile were included to determine if average or slightly above average students could benefit from the Lab. Philadelphia middle grades schools seek to raise their average students' achievement in order to increase their eligibility for one of Philadelphia's competitive-entry high schools.

Schools were given the choice of which grades to include in the study. Eligible students were randomized within grade and regular math class. Randomization within regular math class helped control for math teacher quality. The list of students in each regular math class in the chosen grades was obtained and for each student a coin was flipped. Heads placed students in the Treatment group and they would attend CATAMA in place of a regular elective, such as art or music, and tails placed the student in the

Control group and they would attend another elective. Students were to take the Lab (the experimental group) or the elective (the control group) five days a week, 45 minutes a day, for either 1 semester or 1 trimester (depending upon the school's schedule). A second cycle of students in different grades were then to go through the same process. As discussed in the Implementation Section, these scheduling conditions were not fully met due to implementation problems and school decisions.

All students in the study took a math standardized pre-test at the start of the CATAMA Lab or the other elective and a post-test at the end of the trimester or semester. The growth in math scores between Treatment and Control students is used to determine the impact of the Lab. The Philadelphia School District enacted a common middle grades math curriculum using a single textbook in 2002-03 reducing the possibility that differences in student achievement growth would be due to exposure to different math curricula.

Table 1 describes the study's students' characteristics by their Lab and Control status. Average pre-test scores were almost identical for the two groups. The majority of students are black or Hispanic with the Lab group having statistically significantly fewer Hispanics. The majority of students were in grade 8 with grade 6 providing less than one-third and grades 5 and 7 together 10% of students. School 1 provided half the study's students and School 3 almost 40% with statistically significantly more to the Lab group. Regarding initial math preparation, over one-third of students were on grade, about one-fifth were up to a one and a half grades behind (of these statistically significantly fewer were in the Lab group), less than one-fifth were between 1.5 to 2.5 grades behind, and one-quarter were 2.5 or more grades behind. There was no difference in the percent of students having a regular math teacher with math credentials (secondary or middle grades certified in math)

Table 1 Here

About 5% of students withdrew from their school before the post-test or were absent during the pre or post-test and did not take a make-up. These students differed from the sample in that they were more likely to be female, Hispanic and from School 2. Academically, they had similar levels of math preparation and slightly higher pre-test scores (for the 91% of them that took the pre-test).

Measures and Data Collection

Student math achievement was measured using the CTBS TerraNova math Survey. When the study began in the summer of 2005, the CTBS TerraNova was one of the two District-given standardized tests. As such, it was taken seriously by the schools and students were prompted to take it seriously as well. We have used results from this test in previous research and found it sensitive to school interventions and other key variables linked to achievement such as teacher quality, principal turnover, student mobility, use of NSF-sponsored curricula, and student effort and motivation (e.g., Author, 2002; Author, 2004; Author, 2006). This assessment is designed to measure student achievement from elementary to high school. Philadelphia uses the 2nd edition. Scaled scores provide a continuous measure of student achievement derived using Item Response Theory allowing us to compare growth in achievement among students of different grades (Seltzer, Choi and Thum, 2003).

However, at the start of the school year, the District announced an intention to study whether to continue giving the TerraNova test. In response, we and the schools decided to give the TerraNova math Survey as a pre and post-test as part of the study in order to ensure the availability and comparability of scores. Hopkins personnel administered the test with the teacher in the classroom and carried out make-up testing as necessary. In order to secure students' best efforts, several minutes were spent with each class explaining that the purpose of the test was to evaluate the Lab not the students and that students should try their best so that the school could determine whether or not to maintain the Lab. Hopkins personnel electronically scored the tests and converted them into scale scores using the norms provided by the publisher.

Lab assignment is the treatment (the key independent variable) and is coded as a dummy variable. An additional set of student characteristics was collected from student and teacher records. Student gender, race/ethnicity (Asian, Black, Hispanic and White), and grade level are measured using dummy variables. Student attendance was collected for each cycle and is measured in two ways: 1) the percentage of days attended during a cycle of the study and 2) high attendees (students attending more than the median attendance percentage). Students' initial level of math performance is described by four categories and captured by a set of three dummy variables: 1) on grade - the reference group, 2) moderately underperforming representing .5 to 1.5 school years below grade 3) underperforming representing 1.5 to 2.5 years below grade, and 4) severely underperforming standing for greater than 2.5 years below grade.

We collected two characteristics of the Lab teachers and the students' regular math teachers: 1) years of teaching math and 2) certification (none, elementary, middle school math, or secondary math). Data on the Labs themselves includes: 1) number of days the cycle lasted and 2) number of periods a week the Lab was held for a class. In addition, two dummy variables represent the individual schools (with School 2 the reference group) and capture the unique school conditions affecting student achievement.

Implementation

The impact of the CATAMA Lab, like that any educational reform, depends on its level of implementation (Crandall et al., 1982; Stringfield et al., 1997). To track implementation, we used a weekly observational checklist that noted: 1) availability of all necessary materials, 2) use of the three part teaching routine, 3) level of differentiated instruction, 4) promotion of teamwork, 5) use of motivational practices, and 6) level of student engagement.

Implementation at each school was high in that students attended the Lab for the expected period of time. The use of computerized instruction was also high, although there were interruptions due to hardware problems, with students successfully working in teams at different paces to fill in gaps in their math knowledge. Student engagement appeared high and quickly adapted to the routine of instruction reducing time spent by teachers on classroom management. However, the teacher instructional components (whole class and small group instruction, and motivational practices) were not as well implemented. In part, this was due to the teachers having to learn a different instructional approach. We also found two other contributing factors: 1) overextension of teachers, and 2) teacher self-discipline. Because instruction in the Lab is class and student specific, it requires ongoing preparation by teachers, especially ones new to the Lab. When teachers took on additional educational duties, by their own volition or by school assignment, they lost their preparation time. One Lab teacher was also the school math coach requiring her to work with the entire school's math faculty on a daily basis. Another Lab teacher was given an algebra course to teach. The third was taking an administrator preparation program requiring her to observe other teachers during her preparation time. As a result, the Lab teachers reduced their teaching and motivational activities in favor of more computer instruction allowing them to work on their other job-related requirements or at times to even relax.

School and district technical and policy decisions also had major impacts on the implementation of the Labs. In the late Fall, the district announced that only the 6^{th} and 8^{th} grades' (rather than all middle grades) test scores would be counted toward calculating a school's annual yearly progress that year. In response, School 1 decided to keep the 6^{th} and 8^{th} graders taking the Lab in the Lab after the end of the first cycle and provide them additional computerized test preparation. While these students took their post-test at the proper time, their continued attendance in the Lab for test preparation made it impossible to have a second cycle (that was to focus on 7^{th} grade) at the school. At School 3, the district replaced the principal at the beginning of the school year as well as moved the Lab teacher to a district office position. The new principal, from outside the school, had little knowledge of or interest in the Lab leading to the assignment of a teacher only part-time to the Lab and the use of inferior computers rather than the originally-assigned computer lab. As a result, fewer Lab classes were held during the first cycle and the failure of the computers made a second cycle impossible.

These factors impeding implementation would be typical for any educational program implemented by school personnel and supported by an outside organization especially during the first year of a study in schools (and a district) serving high-poverty high-minority populations.. While the use of all the instructional components were not as high as desired, their low level of use, the expected level of use of the computerized instruction, plus the regular holding of Lab classes for all the treatment students as scheduled combine to give a level of implementation acceptable for studying the impact of the Lab on student achievement.

Analysis and Results

Our analytical strategy addresses the question whether the Lab can effectively enhance middle grades students' math achievement and if so to what degree. We examine the Lab's effect on student gains in math scores. By randomizing students into a Lab and a Control group we control for all observed and unobserved differences at the time of assignment. This randomization also controls for any persistent (time-invariant) effects on learning after the lab assignment. We use bivariate analysis based on a two sample t-test of the mean gains in math test scores of the Lab group versus the Control group to determine if the Lab fosters greater student math achievement. Table 2 shows that, on average, Lab students significantly doubled the gains of Control students. Both Lab and Control students were receiving 90 minutes of math instruction a day in their regular math classes. Lab students received an extra 45 minutes of day of math which was equivalent to an increase of one-half more time of math instruction during the cycle. Even if we scale down the Lab students' gain by one-half, the effect remains high and significant.

To enable the comparison of the impact of the Lab with other educational programs, we calculate an effect size of .26 by standardizing the gain using the standard deviation of the Control group. An effect size of this magnitude is approximately equivalent to the effect of a year of middle school on the mean student gain on math standardized test scores (Bloom, Hill, Black & Lipsey, 2006).

Table 2 Here

In addition to the bivariate analysis, we use a model-based approach to address possible differences that could occur after assignment. For example, we would expect attendance to affect student achievement and attendance occurred after the random assignment. In addition, because the randomization was not made within blocks of individual characteristics, a model controlling for these characteristics can more accurately estimate the treatment effect. These considerations, plus our interest in interaction effects between Lab assignment and student characteristics, justifies controlling for the observed student and teacher characteristics and school indicators through a multivariate analysis

We use an OLS regression model with change scores as the dependent variable (Allison, 1990) to model the effect of the Lab. With 3 teachers and 3 schools, we do not have enough cases for a hierarchical model. However, the inclusion of the dummy variables representing the schools and teachers in our model controls for all unobserved characteristics of the schools and teachers more appropriately than a hierarchical model because it allows for the correlation between these dummy variables and the other regressors, including the treatment. The dependent variable is y_i, the gain in test score for student i. The key independent variable, denoted T_i , takes the value of 1 for the Lab treatment students and O for the control students,. Other control variables measure the characteristics of the students, their regular math teacher and their Lab teacher. However, there were too few regular math teachers and too little variation among them to include their characteristics in the model. Student characteristics (represented by X_i) include: grade level (a dummy variable for 7th & 8th grade), gender, race/ethnicity (Asian, Black, Hispanic, White, Other), initial level of math underperformance (not behind, .5 to 1.5 years behind, 1.5 to 2.5 years behind, and greater then 2.5 years behind), and attendance rate. Differences among the schools are controlled for using a dummy variable for each school (S_i) which also effectively controls for Lab teacher differences because there was only one Lab teacher per school. The model is expressed as:

$$(1) \qquad y_i = \beta_0 + \beta_1 T_i + \beta_3 X_i + \beta_5 S_i + \varepsilon_i$$

All the independent variables, except the treatment, are centered around their specific mean to provide a clearer interpretation. The coefficients for the independent variables remain the same with or without centering. After centering, the intercept, β_0, captures the average gain for the control group and can be interpreted as the gain in test score for the typical student at the mean of each covariate who did not attend the Lab. β_1 captures the average Lab effect in terms of an additional gain for the Lab group when controlling for the other covariates and can be interpreted as the additional gain in test score due to the Lab for the average student. If β_1 is significantly positive and substantial, we have evidence that the Lab successfully increases the Lab students' math achievement as compared to the control group.

Table 3 shows the results from the estimation of the model. A positive significant coefficient of 9.5 was found for Lab attendance. The size of this coefficient is smaller than the 11 points found in the bivariate analysis because we have partialled out any contributions of factors not controlled for by our original randomization of students. The coefficients for gender and race/ethnicity are not significant – an expected result given that they are time-invariant characteristics. Students in the 7th and 8th grades have smaller gains than those in the 5th & 6th grader, similar to the decline in gains as grade increases noted in the literature. Students with the two lowest levels of initial underperformance made significantly greater gains than those with higher initial levels. Higher attendance rates led to marginally significant greater gains. Attendance became significantly positive when the qualitative measure of greater than median attendance was used in place of attendance rate. Neither of the school dummy variables had a significant coefficient. The R^2 was .24 which is relatively high for change models.

Several extensions were made to the model to determine if the Lab had differential impacts on the subgroups defined by student characteristics. Of greatest interest was whether the Lab benefited students as a whole or only at specific levels of initial math performance. Interactions terms between Lab attendance and initial math performance were tested and found insignificant suggesting that the Lab benefits students at all initial math levels studied. Similarly, interactions between Lab attendance and other covariates were also found to be non-significant.

Discussion

Our evaluation of the impacts of the CATAMA Lab has implications not only for the use of the Lab but for policy aiming to increasing math achievement in high-poverty high-minority middle schools. Many students at these schools are performing at such low levels in math that they will require both more and better instruction. Regarding the Lab itself, the results show it to have a clear and sizable impact on student achievement. Lab students doubled the gain of control students. As Lab students spent one-half more time in math instruction during the grading period by attending the Lab, these gains were greater than expected than if students had spent the extra time in their regular class. The Lab appears to provide a more effective form of instruction. When measuring the effect size of these gains, we find that their value of .26 is equivalent to a year of regular math instruction in the middle grades. Using a composite of math standardized tests, Bloom, Hill, Black and Lipsey (2006) found that one year of regular math instruction had an effect size between .19 - .41 for middle grade students (declining as grade increased). In other words, students spending between 30 to 40% of the year for an additional 15-20% of time in math instruction in the Lab make achievement gains equivalent to spending about one year in their regular math class, further evidence that the Lab's instruction is more productive than increasing the amount of regular instruction. Whether these gains are enough to help students better succeed in high school math cannot be answered by this study. However, our future research includes following the 8th grade students into 9th grade to determine if the differences in math achievement continue, and if so, whether at a level of practical importance (such as rates of passing 9th grade math).

In addition, the Lab benefits the variety of students taking part in the study. There were no differential findings by initial level of math achievement, gender, race/ethnicity, attendance, and grade level. As the schools were not randomly selected, we cannot consider the results representative of urban schools serving high-poverty high-minority populations but of only that type of school willing to support a Lab. However, as the students were randomly assigned, we can consider the results representative for the type of students that attends such schools as long as they fall within the eligibility range used by this study. At the schools in the study, one-third to over one-half of students in each class proved eligible to take part.

The relevance of these findings is increased by the practical nature of the Lab. For schools without a Lab, the decision to start one can be made in summer and the Lab can be up and running at the start of the school year. The Lab can reach a large number of students using a medium level of resources. Students can be scheduled into the Lab just as they into other electives. The per-student expenditures for the Lab teacher, her training plus the computers and software are greater than for an additional math teacher teaching 30-35 students at a time but less than the cost for the personnel necessary to run a pull out program serving the same number of students. On average, one Lab teacher can teach five classes of 15-20 students a day reaching 75 to 100 students a semester or 150 – 200 students a year. The Lab can reach a large percentage of students in a school each year while avoiding the interruptions in learning and stigma attached to pulling students from their regular math class to receive special instruction. The Lab also helps with regular math instruction reducing the time math teachers must spend on reviewing more basic concepts and skills. While we were unable to address this point in the study, it is likely the greater gains made by Lab students were due not only to the basic material learned in the Lab but also students' use of this new knowledge to learn in their regular math class.

Our findings can also contribute to policy-making aimed at increasing math achievement at schools serving high-poverty high-minority student populations. Specifically, they suggest that extra help programs should join the list of math reforms (including higher standards, greater accountability, more focused academic curricula, improved teaching, and better learning environments) used to better prepare students for high school math courses. Extra help programs can avoid some of the obstacles that block implementation of other reforms but also help overcome them. These obstacles are often the reason why adding more time for regular classroom instruction may not be as productive as adding extra-help through different forms of organization and instruction.

For example, when teachers are charged with teaching a challenging standards-based curriculum to classes with large numbers of low-performing students, they must usually choose from two non-productive choices. They will either have to teach to the curriculum, even if a substantial number of students cannot keep up because they lack necessary prerequisite skills or understandings, or stop teaching grade-level material and remediate as best they can. Prior experience shows that both choices greatly limit the effectiveness of their efforts to raise students to a Proficient level (Author, 2002). Simply put, if a teacher has to stop grade-level instruction to spend time going over basic fraction concepts or find alternative ways to explain the concept of a variable to a sub-set of students who are struggling with it, they have less time to introduce integers. An effective extra help program closely integrated with classroom instruction provides a third choice. Teachers can depend on the extra-help program to provide students with the more individualized instruction they need to fill in missing knowledge or skills, enabling them to focus on grade level material. If the program is designed to have sufficient capacity to reach most students in need, then classroom instruction can be more effectively accelerated (Author, 1998).

Second, the impact of math reforms is often reduced in high-poverty urban schools by the weak teaching corps. Teachers in such schools are much more likely than those in other schools to lack certification and deep knowledge of content and pedagogy

(Bradley, 2000; Gaskill, 2002; Jerald, 2002; Lankford, Loeb, & Wyckoff, 2002; Monk, 1994; Useem, 2001). Even if new regulations spawned by *No Child Left Behind* solve the basic problem of teacher credentials and content knowledge, (and the latter will apply only to 7th and 8th grade teachers in many states) high-poverty schools will still have to deal with the challenge of high rates of teacher turnover and the induction of many brand new teachers each year (Ingersoll, 2002a, 2002b; Neild and Spiridakis, 2003; Useem, 2003; Useem & Neild, 2002). A strong extra help program can help offset missed learning opportunities when students experience a weak or inexperienced teacher who does not provide strong mathematical instruction.

When considering the implementation of extra help to support math reforms, the CATAMA Lab offers some general guidelines. Organizationally, an adequate extra help program needs to address the majority of eligible students in a manner that ensures they can regularly attend, does not conflict with the school schedule and does not reduce students' on-grade math instruction. The Lab's provision of extra help through a class format (though one of smaller size) during the regular day meets this goal. It fits into the regular school schedule, can include a large number of students through multiple sections, ensures that students will be able to regularly attend (i.e. avoids the difficulties associated with attendance at after school and weekend programs), and does not conflict with students' regular math class. The trade-off is that the Lab substitutes for an elective for part of the year reducing students' exposure to non-academic subjects and generating some student resentment at this loss. The resentment is reduced by scheduling students directly to the Lab so that it is perceived as just another elective class and the opportunity to work with computers.

Instructionally, extra help must address the specific needs of each student and these may differ even among students grouped together in an extra help session due to their having the same relative achievement level. As noted in our results, even students performing above grade-level can benefit from this support. Extra help instruction has to address topics needed to be understood by the whole class, small groups or only individuals. The extra help teacher must be competent not only in teaching the content but also recognizing what gaps the class and individual students have and how to address them on the spot. Computerized instruction also offers a means to address individual needs. By combining the teacher and computer instruction, the Lab offers multiple methods of instruction to address student needs and provides time for the teacher to work with different configurations of students (from whole class to individuals) as the need arises.

References

Abidin, B., & Hartley, J.R. (1998). Developing mathematical problem solving skills. *Journal of Computer Assisted Learning, 14,* 278-291.

Allington, R.L. (1991). Children who find learning to read difficult: School responses to diversity. In E.H. Hiebert (Ed.), *Literacy for a diverse society*: *Perspectives, practices and policies* (pp. 237-252). New York: Teachers College Press.

Bean, R.M., Cooley, W.W., Eichelberger, R.T., Lazar, M.K. and Zigmond, N. (1991). In class or pullout: Effects of setting on the remedial reading program. *Journal of Reading Behavior*, 23(4), 445-464.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years*. Boston: TIMSS Study Center, Boston College.

Bishop J. & Mane F. (2001). The impact of minimum competency exam graduation requirements on college attendance and early labor market success of disadvantaged students. in G. Orfield (Ed.) *Raising standards or raising barriers.* New York: Century Foundation Press.

Blank, R. K., & Langeson, D. (1999). *State indicators of science and mathematics education 1999: State-by-state trends and new indicators from the 1997-98 school year*. Washington, DC: Council of Chief State School Officers.

Bloom, H., Hill, C., Black, A., and Lipsey, M. (2006). Effect sizes in education research: What they are, what they mean and why they're important. Paper presented at the 2006 Research Conference of the Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Bradley, A. (2000). The gatekeeping challenge. *Education Week, 29*(18), 20-26.

Burrill, G. (1998). Changes in your classroom: From the past to the present to the future. *Mathematics Teaching in the Middle School, 4*(3), 184-190.

Cubb J. & Lovelass T. (Eds.) (2002). *Bridging the achievement gap.* Washington DC: Brooking Institute Press.

Crandall, D., Loucks-Horsley, S., Baucher, J., Schmidt, W., Eisman, J., Cox, P., Miles, M., Huberman, A., Taylor, B., Goldberg, J., Shive, G., Thompson, C. & Taylor, J. (1982). *People, policies and practices: Examining the chain of school improvement*, Volumes 1-10. Andover, MA: Network.

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

Fuchs, D., Fuchs, L., Mathes, P., and Simmons, D. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), 171-206.

Gaskill, P.E. (2002). Progress in the certification of middle level personnel. *Middle School Journal*, 33, 33-40.

Ingersoll, R.M. (2002a). The teacher shortage: A case of wrong diagnosis and wrong prescription. *NASSP Bulletin, 86*, 16-30.

Ingersoll, R.M. (2002b). *The teacher shortage: Causes and solutions*. Presentation to the Campaign for Human Capital, School District of Philadelphia. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Jerald, C.D. (2002). *All talk, no action: Putting an end to out-of-field teaching*. Washington, DC: The Education Trust.

Kane, T. (2004). The impact of after-school programs: Interpreting the results of four recent evaluations. Working paper of the William T. Grant Foundation.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Knapp, M.S. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.

Lankford, H., Loeb, S., & Wyckoff, J. (2002.) Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24,* 37-62.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity. *Educational Researcher, 31*(1), 13-25.

Loucks-Horsely, S, Hewson, P., Love, N., and Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mac Iver, D.J. (1991). *Helping students who fall behind: Remedial activities in the middle grades.* (Report #9). Baltimore: Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students.

MacNab, D.S., & FitzSimmons, G. (1999). Enhancing math learning through computer-assisted instruction. *Education Canada*, *39*(1), 38-39.

Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125-145.

Mullis, I., Martin, M., Gonzalez, E., and Chrostowski, S. (2004). *TIMSS 2003 international mathematics report*. Boston: International Association for the Evaluation of Educational Achievement.

National Center for Educational Statistics. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective*. NCES 2005-003, Washington, DC: U.S. Department of Education, Institute of Education Sciences, NCES.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation*. Washington, DC: U.S. Department of Education.

Neild, R. C., & Spiridakis, K. (2003). *Teachers in the School District of Philadelphia: Tables on teacher retention and the distribution of teachers' certification levels and experience in the District by school type, poverty level, and school racial composition*. Philadelphia: Philadelphia Education Fund.

Newmann, F. M., Smith, B. A., 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.

Pelavin, S., & Kane, M. (1990). *Changing the odds: Factors increasing access to college*. New York: College Entrance Examination Board.

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.

Schmidt, W.H., McKnight, C.C., Jakwerth, P.M., Cogan, L.S., Raizen, S.A., Houang, R. T., et al. (1999). *Facing the consequences: Using TIMSS for a closer look at United States mathematics and science education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Seltzer, M., Choi, K., and 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, 263-286.

Stringfield, S., Millsap, M., Herman, R., Yoder, N. Brigham, N., Nesselrodt, P., Schaffer, E., Karweit, N, Levin, M. & Stevens, R. (1997*). Urban and suburban/rural special strategies for educating disadvantaged children: Final report*. Washington, D.C.: U.S. Department of Education.

U. S. Department of Education (1997, October). *Mathematics equals opportunity*. Paper prepared for the U.S. Secretary of Education Richard W. Riley. Washington, DC: Author.

Useem, E. (2001.) New teacher staffing and comprehensive middle school reform: Philadelphia's experience. In V.A. Anfara (Ed.), *The Handbook of Research in Middle Level Education* (pp. 143-160). Greenwich, CT: Information Age Publishing.

Useem, E. (2003, March*). The retention and qualifications of new teachers in Philadelphia's high-poverty middle schools: A three-year cohort study*. Philadelphia: Philadelphia Education Fund.

Useem, E., & Neild, R.C. (2002). *Teacher staffing in the school district of Philadelphia*. Philadelpha: Philadelphia Education Fund.

Wasik, B.A. & Slavin, R.E. (1990). *Preventing early reading failure with one-to-one tutoring: A best evidence synthesis*. Baltimore: Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students.

West, J., Denton, K., & Reaney, L. (2000). *The kindergarten year: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99*. Washington, DC: U.S. Department of Education.

Table 1

Description of Sample Overall and By Control and Lab Groups

Variables	Control	Lab	Total
Pre-Test Scale Score	636	634	635
Male	.49	.41	.44
Female	.51	.59	.56
Asian	.05	.09	.07
Black	.42	.50	.47
Hispanic	.47	.32*	.38
White	.03	.07	.06
Other	.02	.02	.02
Grade 5	.06	.09	.08
Grade 6	.29	.31	.30
Grade 7	.04	.05	.04
Grade 8	.61	.55	.58
School 1	.58	.44	.49
School 2	.15	.11	.13
School 3	.27	.46*	.38
On grade in Math	.31	.37	.35
.5 – 1.5 grades below	.27	.19*	.22
1.5 – 2.5 grades below	.13	.18	.16
> 2.5 grades below	.28	.26	.27
Attendance Rate	.92	.93*	.93
Credentialed Regular Math Teacher	.40	.35	.37
n	172	259	431

* significantly different than the Control Group at $p < .05$.

Table 2

Comparison of Lab and Control Groups' Mean Gains in Math Scale Scores

Lab Group	Control Group	Difference	Effect Size
22**	11	11	.26

** significantly different from the Control Group at $p < .01$.

Table 3

Regression Analysis of Impact of CATAMA Lab on Gains in Students' CTBS Terra Nova Math Scale Scores

Variable	Coefficient
CATAMA Lab	9.5*
Female	-.07
Race/ethnicity (compared to White)	
Asian	.84
Black	6.2
Hispanic	-3.5
Other	-19.6
7^{th} & 8^{th} grade (compared to $5^{th}/6^{th}$)	-14.7**
Initial Underperformance (compared to on grade level)	
.5 to 1.5 years behind	4.2
1.5 to 2.5 years behind	11.7*
2.5 or greater years behind	47**
Attendance rate	.44†
School (compared to School 2)	
School 1	3.9
School 3	3.9

Note: (covariates centered), n = 431, and R^2 = .24.
† $p < .10$.
* $p < .05$.
** $p < .01$.

Appendix 3: AERA Presentation on 9th Grade CATAMA Lab Results

An Extra-Help Math Lab for At-Risk High School Students in 9th Grade Algebra

Allen Ruby (aruby@csos.jhu.edu)
Robert Balfanz
Center for Social Organization of Schools
Johns Hopkins University
March 2008

Abstract

An experimental study of 79 9th grade students at a neighborhood high school in Philadelphia found that students participating in a semester of the Computer and Team Assisted Mathematical Acceleration Laboratory (CATAMA Lab) made significantly larger gains in math achievement than students taking a non-math elective. Lab student gains were 27 points higher than Control students as measured on the CTBS TerraNova Survey Plus standardized math exam. This greater gain represents a difference of over two-thirds of a standard deviation and also a gain of 21 percentiles on a national ranking of 9th graders.

Background

An increasing number of urban districts require students to take algebra during 9th grade, for example, Los Angeles, Portland OR, Baltimore, and Philadelphia. In some cases, this requirement includes passing Algebra 1 for promotion to 10th grade. The goal of this requirement is to increase the amount of challenging coursework taken in high school which has been shown to raise students' academic achievement, foster greater opportunities to attend and succeed in college and provide a wider range of career opportunities (Alexander & Pallas, 1984; Hoffer, Rasinski, & Moore, 1995; Meyer, 1999;

Girotto & Peterson, 1999; Adelman, 1999). These positive impacts have been found for students of all achievement levels ((Gamoran & Hannigan, 2000).

Because historically, low-income and minority students have had less access to challenging mathematics classes (Gamoran & Hannigan, 2000), requiring algebra for all students has been termed a "civil right" (Moses, 2001). By requiring all 9th graders to take Algebra 1, the expected outcome is that all students will have the opportunity to take more advanced math courses in high school. This is important for college as going beyond Algebra 2 has been associated with college entry, avoiding the need for college remediation courses, and college completion. (Adelman, 1999).

However, the policy of placing urban 9th graders in Algebra 1 has led to high failure rates. In 2004, freshman taking Algebra 1 in Los Angeles had a 44% failure rate and only 39% received a grade of C or better (Helfand, 2006). Seven years of data from Milwaukee found an average failure rate of about 50% (Ham and Walker, 1999). One apparent reason for the high failure rates is that the traditional Algebra 1 curse assumes that students have learned middle school math. This assumption may be false in urban districts. For example, Neild and Balfanz (2005) found that only 20% of 8th graders in Philadelphia who went on to attend neighborhood high schools were at grade level on their math standardized tests while over half scored at 6th grade or below. The success of a 9th grade algebra policy for these students includes addressing their gaps in pre-algebra knowledge.

These gaps in 9th graders' math knowledge are primarily in intermediate math knowledge and skills. Students who are below grade level have often mastered basic mathematical operations involving whole numbers, e.g., arithmetic (Campbell, Hombo,

& Mazzeo, 2000). However they may have difficulty using fractions, decimals, percents, and negative numbers (Kilpatrick, Swafford, & Findell, 2001) in part because not all middle grade students receive effective instruction in them (Mullis et al., 2001; Cogan, Schmidt, and Wiley, 2001). These studies using the TIMSS also found a lack of exposure to other advanced math topics that may be assumed in Algebra 1 courses including proportional reasoning, probability, measurement and geometry.

Requiring all 9th graders to take Algebra 1 without addressing their middle grade math gaps can lead to several types of failure. First, the students themselves may fail the course. Second, a high level of student failures could lead to either dropping the requirement that all students take Algebra 1 or reducing the demands of an Algebra 1 course (both of which can potentially reduce opportunities for students). An alternative approach is to provide a means to fill student math gaps without reducing the requirements of the Algebra 1 course.

The Intervention: The CATAMA Lab

The Computer and Team Assisted Mathematical Acceleration Laboratory (CATAMA Lab) is an elective course for students needing additional assistance in math while they continue in their regular math class. The Lab helps students fill in gaps in math skills and knowledge that they are incorrectly presumed to have already learned in earlier grades and also can be used to preview upcoming material from the regular math class. Class size is reduced and students attend the Lab for about one semester in place of an elective course (such as art or music).

The Lab is taught by a full time, certified, and experienced mathematics teacher. The Lab teacher receives an initial day of professional development and weekly in-class support provided by a Lab trainer with experience in both teaching the Lab and supporting Lab teachers. Typically the teacher instructs five sections of 15 to 18 students per day. Each class is taught using three main instructional components. The mix of instructional methods helps maintain student interest, offers students different ways to learn the material, and provides individual students with instruction geared to their needs (both through computer instruction and teacher tutoring)..

Class begins with approximately 15 minutes of whole group instruction on skills and concepts students that students are known to lack and will be required to use in their regular math class. This both helps the students learn the concepts and it helps them stay interested and focused in their regular math class rather than becoming frustrated by their lack of comprehension and giving up.

Class continues with 20-30 minutes of individualized computer and peer-assisted instruction. Each lab has 10 to 15 networked computers. Students typically spend between 20 and 30 minutes per day on the computers using instructional software tailored to their needs. To address gaps in middle grades math, students work with Larson's pre-algebra software. This software includes formative testing to determine what concepts a student has not mastered, instruction in those concepts, and then summative testing to determine if a student has learned the material.

Students are paired and then teamed with students at similar skill levels. Peer-assisted learning techniques are taught so that the students learn to work together though they take the tests individually. Working in teams helps students stay focused on the work, motivated to keep going, and take the time to discuss the problem rather than rush to attempt a solution.

The computer and peer-assisted learning features of the Lab also provide the teacher with the time for the third instructional component of the class - individual or small group tutoring. While most of the class is working on the computers, the teacher can provide direct tutoring to individual or small groups of students. As students enter high school with different gaps in their math skills, this time allows the teacher to address individual student needs without holding up learning for the rest of the class.

Providing extra help in math through the Lab has several practical benefits. First, unlike pull out programs, the Lab does not interfere with student attendance to their Algebra 1 class. Second, the Lab takes place during the school day avoiding the low attendance problems affecting after-school/Saturday and summer school programs. Third, the Lab allows math remediation to be done outside the regular math class so that the Algebra 1 teacher can focus on teaching algebra.

Study Design

Seventy-nine 9th grade students taking algebra in 2006-07 in a Philadelphia neighborhood high school were randomly assigned to either a CATAMA Lab (48 students) or to a non-math elective class (31 students) for 63 full school days during first semester. The comparison is then CATAMA Lab versus things as they are normally. Control students are not receiving extra math. This is an efficacy study to determine whether the Lab has a positive effect on students' math achievement.

Assignment was made within their regular algebra class. There were five freshman 9th grade Algebra 1 classes taught by two teachers (one had two sections and the other three sections) using the same textbook and pacing guide. Students eligible for the study scored in the mid-range (25th to 70th percentile) on their 8th grade standardized math test. Previous work in the district's schools led to a finding that students scoring below the 25th percentile needed individual tutoring and/or additional services to succeed. Students attended the Lab or the elective every day for 1 class period while continuing with their daily algebra class.

Students in the Lab received teacher and computer instruction in eight math modules including: percents, geometry in a plane, ratios, rates and proportions, coordinate geometry, probability, algebraic expressions, and algebraic equations. Students moved at different paces through these modules and as result not all completed the final two. Where students needed additional assistance, the teacher provided class and small group instruction on more basic math topics, for example, order of operations, fractions, decimals, and positive/negative numbers. This curriculum was developed to cover some of the standards for the 9th grade while also providing a heavy emphasis on areas where students as a whole scored low on the previous year's standardized test

All students in the study took a math standardized pre-test at the start of the CATAMA Lab or the other elective and a post-test at the end of the semester. By comparing the growth in math scores (from pre to post-test) between Treatment and Control students, we will determine the Lab's impact on math achievement. The test used to measure achievement is the Comprehensive Test of Basic Skills (CTBS) TerraNova Mathematics Survey Level 19, Form A. The assessment is a standardized norm-referenced achievement test with versions for grades 2 to 12 published by CTB/McGraw Hill. The test is not focused on algebra, though it contains several algebraic items, and so does not measure how much algebra students learned in their regular 9th grade math class. It was chosen because it tests a broad range of math skill often found lacking among the type of 9th graders examined in this study. This lack was the impetus behind the use of the CATAMA Lab. Students took a pre-test on September 19th with make up exams given the rest of the week. The post-test was on January 23, 2007 with make up exams held the rest of that week. The tests were given by Hopkins personnel during algebra class with the math teacher in attendance.

Comparison of Lab and Control Students

The randomization of students into the Lab and control groups should ensure that the treatment and control students were similar to begin with. Randomizing with each algebra class also controls for differences in the type of instruction provided by the two teachers. In addition, comparing growth in test scores will help control for non-observed factors that might have been unequally distributed due to unfortunate randomization that affect test scores. Table 1 compares the two groups, specifically their initial test scores, the proportion of gender and race/ethnic groups making up each group, and the grade level equivalent of the students based on their pre-test. An asterisk by a Lab student value means that there it is statistically significantly different from the value for the Control students.

Looking at pre-test scores, we see that on average Lab students scored 6 points higher than Control students but this was not a significant difference. The only statistically significant difference between the two groups is that the Lab group contained a smaller percentage of black students than the Control students. On every other measure, there are no significant differences, i.e., the groups were statistically similar before the experiment started.

Table 1: Comparison of Lab and Control Groups

Variables	Lab Students	Control Students	All Students
Pre-Test Scale Score	659	653	657
Male	.56	.45	.52
Female	.44	.55	.48
Asian	.19	.10	.15
Black	.52*	.74	.61
Hispanic	.13	.10	.11
White	.17	.06	.13
Algebra Teacher 1	.42	.45	.43
Algebra Teacher 2	.58	.55	.57
On grade in Math	.25	.23	.24
1 grade below	.27	.19	.24
2-3 grades below	.19	.19	.19
> 3 grades below	.29	.39	.33
n	48	31	79

* significantly different from Control students at .05 level

Results

We examine the Lab's effect on student gains in math scores by examining student gains between the pre-test and the post-test and also by comparing student math grades. We check to see if Lab students made greater gains than Control students and use a two sample t-test of the mean gains to determine if any difference is statistically significant. If it is, we have evidence to support the hypothesis that the Lab fosters greater student math achievement. Table 2 shows the results of this analysis. It shows that, on average, Lab students significantly outgained Control students by 27 points – equal to almost two-thirds of a standard deviation in gains.

Table 2: **Comparison of Lab and Control Groups' Mean Gains in Math Scale Scores**

Lab Group	Control Group	Difference	Effect Size
29**	2	27	.63

** significantly different from the Control Group at .01 level n = 62 students

Another way to think about these results is to compare students' ranking on the national performance of 9th graders on this test. Table 3 shows what percentile the two groups were ranked on the pre-test and how this ranking changed on the post-test. We see that both groups were ranked similarly on the pre-test with Lab students performing at the 33rd percentile on average (they performed better than 1/3 of the students around the country who took this test but worse than 2/3) and Control students performing at the 31st percentile. However, the two groups' rankings varied widely on the post-test. The Lab group rose 17 percentiles to the 50th percentile (they performed at the median) while the Control group actually dropped 4 percentiles in the national rank. As a result of this drop, Lab students gained 21 percentiles more than Control students.

Table 3: Comparison of Lab and Control Group Percentile Rankings on Math Test

	Pre-Test Percentile Rank	Post-Test Percentile Rank
Lab students	33	50
Control Students	31	27

We also compared student math grades. As the goal of the Lab is to help students succeed in Algebra 1, grades are a key outcome. While grading may differ by teacher, this study only includes two teachers and randomized students with their classes reduces the impact of differences in grading. These grades were submitted by the teachers the same week that post-testing was done (and received by students two weeks later) so they fully reflect any impact the Lab may have had on students' performance in their Algebra 1 class. Table 4 shows that Lab students had a larger percentage of A grades while

Control students had a larger percentage of D grades. About one-third of both groups had failing grades.

Table 4: Percent Distribution of Math Grades

Mid-term Grade	Lab Students	Control Students	All Students
A	28%	8%	19%
B	14%	15%	15%
C	19%	23%	21%
D	6%	19%	12%
F	33%	35%	34%

Sensitivity Analyses

There are two potential concerns with the positive findings discussed above for the CATAMA Lab: 1) 22% of the students dropped out of the study, and 2) 19% showed negative gains on the post-test. In this section, we examine the importance of these two factors.

A. Study Dropouts

Our original sample had 79 students but only 62 students completed the study by taking the post-test. Of the 17 students who dropped out of the study, 12 were assigned to the Lab (25% of the original Lab group) and 5 to the Control group (16% of the original Control group). Dropping out of the study occurred through several processes: 1) the major process was by students withdrawing from the school, 2) students did not attend school for the week of the post-test and the make-up tests, or 3) students were at school but did not take the test seriously – they refused to take it or drew on the answer sheet. If these dropouts were poorly performing students, than the results for the Lab might be biased since a greater proportion of Lab students dropped out than Control students.

Table 5 compares the Lab dropouts with the Control dropouts to check if they differ. There are no statistical differences between the two groups. Because of the small number of them this is not unexpected. The differences between the two might seem to favor the Control group as the Control dropouts had a lower mean test score and a larger percent of students who began the study three or more grades below level.

Table 5: Comparison of Dropouts from Lab and Control Groups

Variables	Lab Dropouts	Control Dropouts
Pre-Test Scale Score	647	631
Male	.83	.40
Female	.17	.60
Asian	0	0
Black	.75	.60
Hispanic	.08	.40
White	.17	0
Algebra Teacher 1	.50	.60
Algebra Teacher 2	.50	.40
On grade in Math	.17	0
1 grade below	.33	.20
2-3 grades below	.17	.20
> 3 grades below	.33	.60
n	12	5

A second way to examine the impact of the dropouts is to compare the remaining 62 Lab students versus control students to examine if there are any significant differences between them. Table 6 shows this comparison. As in the original comparison (Table 1) only the proportion of blacks in the Lab group is significantly different than in the Control group. There are no significant differences in the other variables including the pre-test score.

Table 6: Comparison of Lab and Control Groups

Variables	Lab Students	Control Students	All Students
Pre-Test Scale Score	662.8	657.5	660.6
Male	.47	.46	.47
Female	.53	.54	.53
Asian	.25	.12	.19
Black	.44*	.77	.58
Hispanic	.14	.04	.10
White	.17	.08	.13
Algebra Teacher 1	.39	.42	.40
Algebra Teacher 2	.61	.58	.60
On grade in Math	.28	.27	.27
1 grade below	.25	.19	.23
2-3 grades below	.19	.19	.19
> 3 grades below	.28	.35	.31
Attendance Above the Median	.47	.54	.50
n	36	26	62

* significantly different from Control students at .the 05 level.

We take one more step to ensure that the student attrition did not overly change the composition of our two groups. We estimate a logit model for students who withdrew versus students who did not using our independent variables from Table 1. This model estimates the odds of a student withdrawing given their characteristics (e.g. Lab enrollment, race/ethnicity, gender, pre-test score, etc.). If the coefficients on any of the independent variables are significant, this will provide evidence that our two groups now differ on this variable. For example, if the coefficient on pre-test is positive and significant, this is evidence that students who scored lower on the pre-test were more likely to withdraw raising the possibility that our two groups are no longer similar on pre-test scores. Because we have a small sample size, we collapse some of our independent variables. The four race/ethnicity variables become either black and non-black or black/Hispanic and Asian/white. The four grade performance variables become on-grade and below grade. None of the coefficients from the logit model are significant so we do not report them here. As the coefficient for the Lab was also not significant, we have no evidence that enrollment in the Lab increased or decreased the odds of dropping out of the study.

Based on these three comparisons, the loss of 17 students does not appear to have significantly changed the composition of the two groups on the student characteristics we are able to observe.

B. Negative Test Gains

Of the 62 students available for study, 15 actually lost ground on the post-test and had negative gains. Of these, 5 were Lab students and 10 were Control students. Overall, 5 students had major declines (over 49 points) and 4 of these were Control students. There are three ways to view this outcome. First, that it is normal - students can do worse on a post-test because they have forgotten material, become confused by new material learned, or lose interest in taking the test. Most studies make this assumption and randomization of students ensures that there will be an equal probability of such students being in both groups.

Second, this result can be interpreted as further evidence that the Lab has a positive effect on student achievement. Fewer Lab students become confused about material they already knew and/or the Lab motivated them to do well on the test. Note that Lab and Control students from the same algebra class took the pre and post-tests together so they received the same encouragement at testing time to do well.

Third, the randomization may not have successfully distributed students with a tendency to do worse on the post-test or the attrition of students may have led to more such students remaining in the Control group. Because more Control students had major losses on the post-test, we are concerned that this might skew the results in favor of the Lab students. To test the importance of these negative gains we redo the test of the significance of the differences in the average test score for the Lab versus the Control students without those students who had major losses (-49 or more points). As a second test, we also drop students who had large losses (-29 or more points). These were natural cutpoints in the data: 3 Control and 1 Lab student had losses of -49 or more points and 5 Control and 1 Lab student had losses of -29 or more.

Table 7 shows the results of the tests. In both cases, Lab students continue to make significantly larger gains than Control students (21 points and 16.8 points respectively with the latter having a reduced effect size of .38). As expected, these gains are smaller than the original test which found a difference of 27 points.

Table 7: Adjusting for Negative Gains

	Lab Students Scaled Score	# of Lab Students	Control Students Scaled Score	# of Control Students
Test 1	31.8*	35	10.8	23
Test 2	31.8*	35	15.0	21

* significantly different from Control students at .05 level.

Dropping those students with large negative gains on their post-test also improved the Control group's change in national percentiles. The Control group's ranking rose 1 percentile after dropping students with losses of -49 or more points and 3 percentiles after dropping students with losses of -29 or more points. The Lab group maintained its gain of 17 percentiles in both cases. The difference in percentile gains between the two groups though somewhat smaller than at first (21 percentiles as shown in Table 3) remains large (14-16 percentiles).

Discussion and Future Research

The results show the Lab to have a clear and sizable impact on student achievement. Lab students made large gains in test scores and national rankings while Control students made small score gains and actually dropped in national rankings. Fewer Lab students also had net losses in test scores and the Lab's success continued even when adjusting for these losses in the Control group. Lab students also had, on average, higher Algebra 1 grades than Control students. However the Lab had no obvious effect on preventing math failure as about one-third of students in both the Lab and Control groups had failing grades. Lab students spent double the time in math instruction during the grading period they attended the Lab but showed far more than double the gains than expected than if students had spent the extra time in their regular math class.

The next step in evaluating the Lab is to compare its impacts versus those of alternative approaches of providing extra-math instruction in middle grades materials such as after-school programs (including summer school) or in-school alternatives (such as extended class time or computer instruction that does not include the other instructional components of the Lab). This work is necessary to determine the effectiveness of the Lab and whether resources would be better invested in the Lab or some alternative. An additional research topic would be the impact of including more Algebra 1 materials in the Lab to determine if these would increase student success in Algebra 1. Linked to this work, would be qualitative research on why students are failing

Algebra 1 to check if academics are the key reason or some other services need to be combined with the Lab to raise student success in Algebra 1.