Professional Development and the Achievement Gap in Community School District #2

Michael Harwell, Laura D'Amico, Mary Kay Stein,
and Guido Gatti
Learning Research and Development Center
University of Pittsburgh

February 15, 2000

This paper was prepared under the sponsorship of the High Performance Learning Communities Project at the Learning Research and Development Center, University of Pittsburgh, under

research contract #RC-96-137002 with the Office of Educational Research and Improvement at the U.S. Department of Education. It is the Student Achievement deliverable #0014 (4.5.c).

Professional Development and the Achievement Gap in Community School District #2

Michael Harwell, Laura D'Amico, Mary Kay Stein, and Guido Gatti
Learning Research and Development Center
University of Pittsburgh

February 15, 2000

Teacher professional development is seen as a possible means of improving the likelihood that all *students will meet high academic standards. To test this assumption, we review student achievement in mathematics and literacy for students in a highly diverse urban district—New York City's Community School District #2. Variations in achievement patterns for District #2 students of different ethnic, linguistic, and socio-economic backgrounds are described. Teacher responses on a questionnaire about their engagement in professional development activities are then summarized. Finally, a multilevel data-analytic model is used to explore the effects of teacher engagement in professional development on student achievement. The results of these analyses reveal little influence of professional development on achievement in either mathematics or reading in District #2. However, due to various data collection and sampling difficulties, we view these findings as inconclusive.*

America's urban districts are in distress. Supporting children's learning is more demanding and more essential than ever before, yet teachers who work in urban districts often find themselves engulfed by archaic rules, unbending regulations, and rigid bureaucratic structures. Moreover, because of the chronic underfunding and difficult working conditions, urban schools generally attract only the least experienced and often the most poorly trained teachers. Although there are examples of classrooms and even whole schools who serve urban children well, making high-quality education the rule rather than the exception in American cities is perhaps the biggest challenge facing public education today.

One response to this challenge has been increased attention to professional development for all teachers, but especially for those who work in inner city schools. Advocates of strong and continuous professional development suggest that, done well, it will improve teachers' skills, confidence and knowledge, thereby developing the capacity of schools to deliver quality instruction. Better instruction in turn will lead to more (ideally, *all*) students achieving high academic standards.

Despite consensus regarding the need for more and better professional development, there are few research findings that point unambiguously to the fact that professional development does indeed improve student achievement or to the kinds of professional development that are most likely to lift the achievement of all students. Indeed, only a small fraction of studies on teacher professional development even include measures of student learning (Kennedy, 1998). Those studies which do explore the impact on student achievement, however, present a strong case “for attending more to the content of inservice teacher education and for attending less to its structural and organizational features” (Kennedy, 1998, p. 16-15). In the studies reviewed by Kennedy, those professional development experiences which focused on teachers’ knowledge of the subject, on the curriculum, or on how students learn the curriculum, demonstrated the largest influences on student learning.

The educators of Community School District #2 in New York City have based more than a decade of reform on assumptions such as these. First, they’ve shaped their efforts at district-wide instructional improvement through subject-matter initiatives. For example, after investing considerable time and energy researching how students learn to become literate, district professionals identified and adapted an elementary literacy program that, in their judgment, would support student learning in ways consonant with the research. Then they built a professional development system that would enable teachers to enact the program successfully. Although this professional development system can be described in terms of the wide array of structural and organizational forms that it encompasses,[1] we’ve argued elsewhere that its most important feature is the manner in which it parallels and supports the kinds of learning demanded of students in the classroom (Stein, D’Amico & Johnstone, 1999).

There is some indication that District #2’s professional development strategy has been a successful one. Between 1988 and 1998, the percent of students achieving at or above grade level rose from 56% to 73% in reading and 66% to 82% in mathematics (HPLC Technical Report). (See Figure 1.)

Nevertheless, it is difficult to empirically tease out the factors that theoretically could be responsible for this rise in student achievement. Over the years, District #2 has leveraged a variety of tools and resources to help underachieving schools and students. For example, the district maintains high standards for hiring teachers and principals; they are also vigilant about maintaining high standards of practice. In addition, District #2 carefully targets its resources at those students who need the most attention (see Stein, Harwell, & D’Amico, 1999). They provide extended day and extended year programs for students that attend schools with a history of low achievement, reading recovery teachers for at-risk readers, and “push-in” teachers for classrooms with a high proportion of children reading below grade level. Finally, high poverty schools in the district maintain school lunch and breakfast programs, as well as medical and social services that help support the basic needs of children without

[1] The forms range from workshops to staff developers observing and co-teaching with individual teachers in their classrooms to various kinds of intervisitation and co-teaching between teachers (Elmore & Burney, 1996).

which learning is difficult at best. All of these forms of assistance, a subset of them, or additional unidentified factors may be responsible for the persistent rise in student achievement scores.

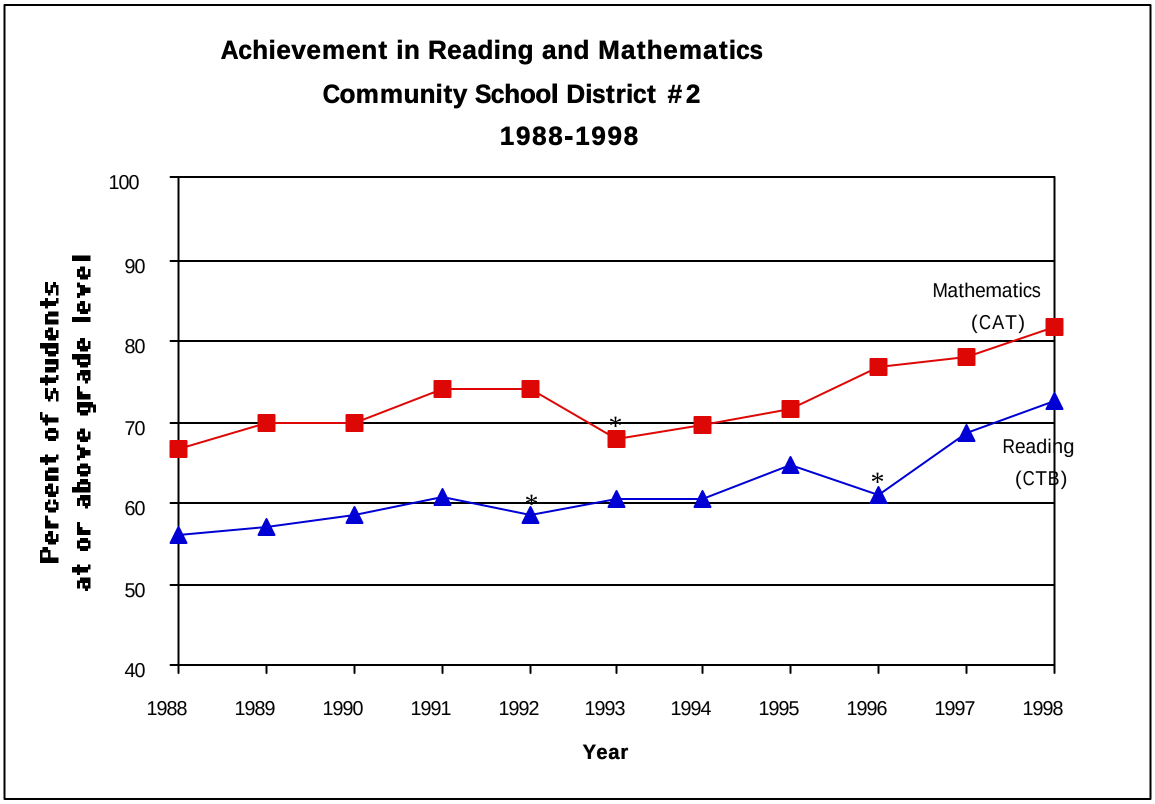

Figure 1: Reading and mathematics achievement in District #2 between 1988 and 1998 as measured by the Comprehensive Test for Basic Skills (CTB) and the California Achievement Test (CAT). (*) Indicates a test renorming.[2]

The purpose of this study is to better understand the influence of a variety of factors on student achievement in District #2 with a particular interest in the role of teacher professional development. Previous attempts to understand the role of professional development in District #2 have been conducted using indirect measures and with the school as the unit of analysis. One study (Resnick & Harwell, 1998) suggested that the quality of professional development in schools (as judged by the Deputy Superintendent and the Director of Professional Development) was positively related to the achievement of students in those schools. In this study, we use teachers' self-reports of their engagement in various kinds of professional development activities as our measure of teacher professional development. We measure student achievement by individual student performance on the Comprehensive Test of Basic Skills (CTB, 1991) in both reading and mathematics.

[2]Data obtained from District #2 archives. According to the notes made on these records, the scores are based on all students in 1988 and General Education and Reading Resource rooms for 1989-'96. The reading scores were renormed in 1992 and 1996 and the mathematics scores were renormed in 1993.

We begin by describing the characteristics of students in District #2 in terms of their socio-economic status, ethnicity, English proficiency, gender, and their achievement in mathematics and reading. We then describe the challenges that the district faces in closing the achievement gap between advantaged students and those who come to school poorly prepared to learn. In the final section we ask the question, "Are teachers with strong professional development participation patterns more likely to have closed the achievement gaps?"

METHODS

Student data

The study was conducted during the 1998-99 school year. Achievement and demographic data for individual students in District #2 was made available through the Division of Assessment and Accountability of the Office of the Deputy Chancellor of Instruction of the New York City Board of Education. The information obtained included the ethnicity and gender for every student, their eligibility for free or reduced priced lunch, their English proficiency status, their attendance rates, their achievement scores in reading and mathematics, and their enrollment status for three years (1996-97, 1997-98 and 1998-99).

District #2 is one of New York's Community School Districts and thus services primarily elementary and middle school students. Of the approximately 22,600 students enrolled in the 1998-99 school year 12,045 (53%) were in kindergarten through fifth grade, and 6,224 (28%) where in grades six through eight. The 49 schools are configured in a variety of ways, the most common of which are PK/K-5 (20), 6-8 (8), PK/K-8 (4), and 6-12 (4). The size of these schools varies rather dramatically, from less than 100 students in some of the specialty option schools, to more than 1000 in comprehensive middle schools. Because achievement testing does not begin until the third grade and because high school age students are a small portion of the entire population, our study is limited to the 13,237 students enrolled in grades three through eight.

Students in District #2 take a number of tests mandated either by the district, the New York City Board of Education, or the State of New York. Many of these are given to students in specific grades. For example, in 1998-99 the State of New York introduced new performance assessments in literacy and mathematics that are administered to students in grades four and eight. In the 1998-1999, the CTB in mathematics and reading were the only tests given to all students in grades three through eight and we use them as our measure of achievement. These tests were taken by the students in the spring of 1999. Complete data for achievement scores and demographics was available for 11,339 of the 13,327 students in grades three through eight.

Teacher data

Information on teachers' backgrounds and professional development experience was obtained through a questionnaire sent to District #2 teachers in June of 1999. (See Appendix A.) The questionnaire had several sections, including: education and certification; gender and ethnicity; teaching experience (number of years, which grades, etc.); engagement in the professional community of their school (e.g. "How often in the past academic year did you have detailed discussions about instructional practice in mathematics with other teachers in your school?"); engagement in the professional community of District #2 (e.g. "In the past academic year did you serve as a mathematics lead teacher?"); and engagement in the professional community beyond District #2 (e.g., "In the last two years did you attend a workshop or conference outside your district?").

The limitations of self-report measures are well known (e.g. Gay, 1996; Stein & Henningsen, 1992). Those most pertinent to this study are the possibility that teachers will respond in socially desirable ways (e.g., some teachers may not feel comfortable indicating they *never* have substantive conversations about instruction with their colleagues), or have difficulty accurately estimating the frequency with which they engage in the various professional development activities described.

The questionnaires were sent to school principals about two weeks before the end of school with the request that they be distributed to all teachers with students in grades three through eight. 99 completed questionnaires were received by September 1999. 37 of these questionnaires were unusable because they came from teachers in grades other than three through eight. The remaining sample of 62 questionnaires represent approximately 12% of District #2's third through eighth grade teachers. These teachers have a total of 1,536 students, or 12% of the students in grades three through eight. Some of the initial data analyses on the achievement patterns of students from different socio-economic, ethnic and linguistic backgrounds are done for all students in grades three through eight. Those analyses that link achievement to professional development, however, are restricted to this smaller sample of 1,536 students.

Analyses

A shortcoming shared by previous research done on the effectiveness of District #2's professional development system (Resnick & Harwell, 1998; Stein, Harwell & D'Amico, 1999) was that the units of analysis used in these studies were schools. As a result, variation among students' performance and teachers' experiences *within* schools was ignored. A defining feature of American public education is that educational organizations have a hierarchical structure. For example, students are nested within classrooms run by particular teachers, who are nested within schools. Incorporating this multi-level structure into the data analysis permits a more realistic approach to understanding the predictors of student achievement. Multilevel data-analytic models provide a mechanism for statistically modeling the effects of this hierarchical structure that avoid the problems of aggregation bias and misestimation of standard errors of parameter estimates that have plagued educational research (Bryk & Raudenbush, 1992). The need for multilevel models in school-based research is now widely acknowledged, and numerous examples of these models can be found in the educational research literature (e.g. Grolnick, Benjet, Kurowski & Apostoleris, 1997; Lee & Bryk, 1989; Newmann & Associates, 1997; Raudenbush, Rowan & Cheong, 1993; Rumberger, 1995; Sui-chu & Williams, 1997).

Harwell and Gatti (1999) addressed the units of analysis shortcoming of the Resnick and Harwell and Stein, et al. studies by performing a multilevel analysis that examined variation in student achievement within and between schools in the district. However, there was no information available at that time to connect student performance to particular classrooms and/or teachers and the effect of teachers' engagement in professional development on achievement was not studied.

For the current study, we describe the characteristics and achievement patterns of the District #2 student population overall in grades three through eight and in the sample of 1,536 students. We then use regression models/hierarchical linear models (HLM) (Bryk & Raudenbush, 1992) to explore the extent to which teacher engagement in professional development activities mediates achievement gaps between students of different socio-economic, linguistic and ethnic backgrounds. We also review the extent to which student participation in District #2 classrooms (as measured by their attendance in school and the number of years enrolled in District #2) is related to higher achievement. Finally, since district leaders made a conscious decision to prioritize professional development over class size, we explore the effect of class size on achievement.

Data issues

The conclusions that can be made from these analyses are limited in several ways by problems associated with the available data. In addition to the issues associated with the use of teachers' self-reports of activity mentioned earlier, there are problems with missing data, sample reduction due to lack of variation in some of the data, and the links between achievement data of students in grades 6-8 with their teachers. (See Appendix B for details.) These various problems lead to analyses which may not adequately represent the achievement of Asian students (since many did not take the tests in mathematics and reading) and which do not include classrooms which are homogenous with respect to the ethnicity or socio-economic status of the students. Despite these problems, we believe our findings can provide useful initial insights into the effect of teachers' engagement in professional development on student achievement in District #2.

STUDENTS AND TEACHERS IN DISTRICT #2

The schools of District #2 reflect the extraordinary diversity of Manhattan. Some schools are in the West Side's Hell's Kitchen, and some in extremely affluent areas on the Upper East Side (Elmore & Burney, 1998). The student body as a whole and the students in our sample mirror this diversity. (See Table 1.) In grades 3-8, half the students are male, a little over a third of the student population is Asian, a little under a third is White, about 20% is Hispanic, and approximately 14% is Black. A significant proportion of the student population includes new immigrants who may be more likely to be classified as English Language Learners. Overall, about 12% of the students in grades three through eight are classified as English Language Learners and therefore eligible for special services and in some cases exempt from testing. 74% of those not yet proficient in English are Asian.

	Grades 3 through 5		**Grades 6 through 8**	
	All Students	Sample	All Students	Sample
Eligible for free/reduced lunch	60%	60%	59%	40%
English Language Learners	13%	5%	11%	9%
Ethnicity				
Asian	36%	31%	36%	35%
White	31%	39%	28%	25%
Hispanic	21%	14%	19%	24%
Black	13%	16%	15%	15%
Gender				
Male	50%	51%	50%	47%
Female	50%	49%	50%	53%

Table 1: Demographics for District #2 students in grades 3-8 during 1998-99

Eligibility for free and reduced lunch is often used as an indicator of socio-economic status in educational research (Ensiminger & Slusarcick, 1992; Entwisle & Alexander, 1992). Using this measure, District #2 is a fairly wealthy urban district. With approximately 60% of their students eligible for free or reduced lunch, they rank as the fourth wealthiest community school district in New York City (HPLC Technical Report I) and are in the upper quartile for urban districts nationally[3] (Council of the Great City Schools, 1999). At the same time, they are one of the most socio-economically diverse districts in New York City. Schools have anywhere between 100% and less than 12% of their students

[3]Data obtained from the Council of the Great City Schools Website (www.cgcs.org). Most of their data is from 1997-98, but for some districts it is from 1996-97.

eligible for free or reduced price lunch. (See Figure 2.) This means that professional development in the district must serve the needs of teachers in both high-poverty and affluent school settings.

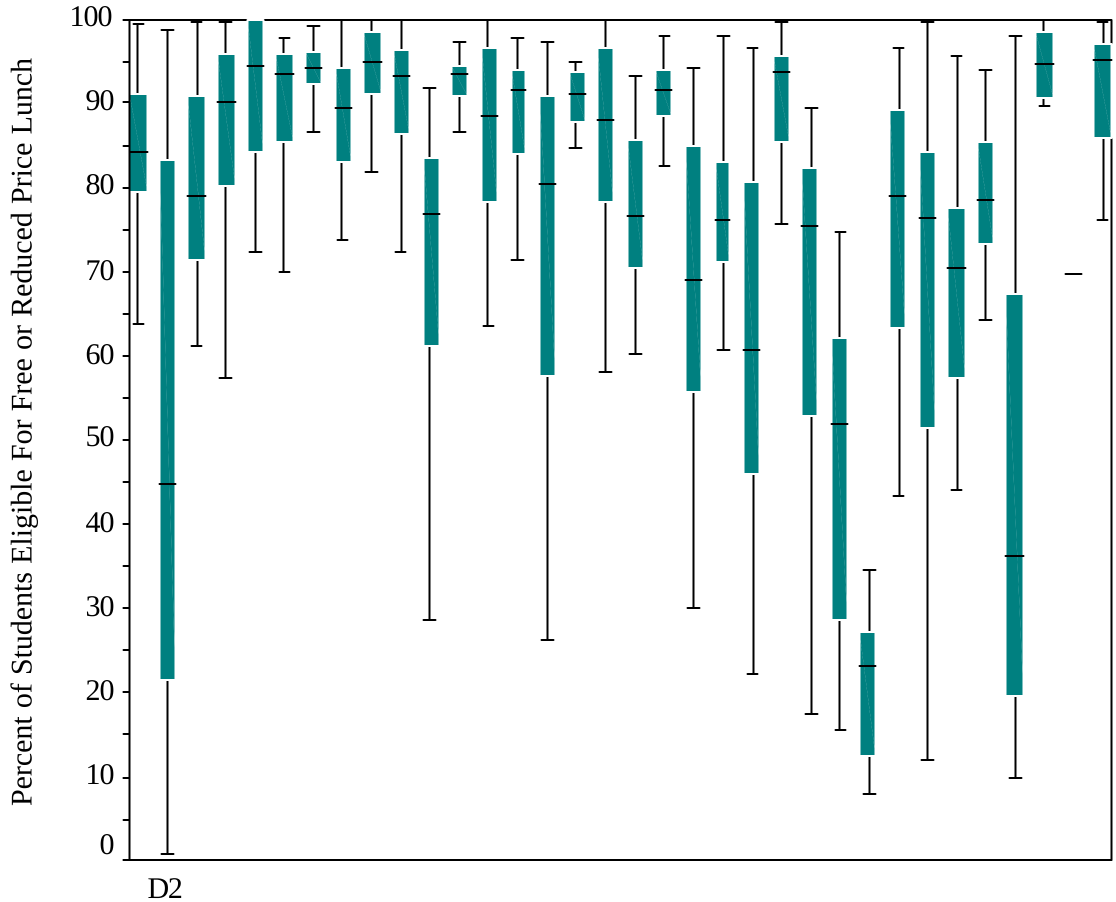

Figure 2: Variation in the percent of students eligible for free or reduced price lunch per school in New York City's community school districts[4]

On the whole, the students in the sample of 62 classrooms for which teacher data are available are similar in Ethnicity, English proficiency and gender to the rest of the students in grades 3-8. There are slightly more males in the sample than in the district as a whole and the rate of English proficiency is a bit higher. The only difference of any magnitude between the sample and the rest of the students in grades 3-8 are the number of middle school students eligible for free or reduced price lunch. In general, the teachers for the grade 6-8 sample have significantly fewer impoverished students (40% eligible for free or reduced price lunch) than the teachers of middle school age students in the district as a whole (59%).

[4]Data obtained from School Report Card database made available by New York City's Board of Education (www.nycenet.edu) and is available on-line at http://207.127.202.63.

Student achievement

The average achievement in both mathematics and reading as measured by the CTB was slightly lower for those students in the sample than the rest of the students in grades 3-8. (See Table 2.) Students in the sample scored on average 11.4 points lower on the mathematics exam and 8.02 points lower on the reading exam. These differences were small and not statistically significant at the $\alpha = .05$ level. However, the variances of the two groups did differ, with the achievement scores of the sample varying significantly less than those of the rest of the students. Thus, the test performance of the sample students is more homogenous.

	Students not in sample Grades 3-8	Sample students Grades 3-8	Difference
Average mathematics score	696.21	684.87	11.4
Average reading score	691.42	683.4	8.02

Table 2: Average achievement of District #2 students in grades three through eight on the 1999 CTB in mathematics and reading

Attendance and Time in District #2

As shown in table 3, most of the grade 3-8 students (88%) had been enrolled in District #2 schools for at least three years.[5] 10% had been enrolled for two years and only 1% had been enrolled for just one year. Attendance among the students in grades 3-8 was high and did not show much variation. On average students were in class for 94% of the days during which class was in session (n=13,236, SD = 5.96). The sample showed very similar attendance and enrollment patterns.

	All students Grades 3-8	Sample students Grades 3-8
Average Percent of Days in Attendance	94%	94%
Time in District #2		
1 year	1%	1%
2 years	10%	7%
3 years	88%	92%

Table 3: Attendance and time in District #2 rates

[5]Students may have been enrolled in District #2 schools for a longer period, but we only have enrollment data going back three years.

Teachers

Overall, the 62 teachers from whom we received usable questionnaires tended to be white (77%) and female (82%). Almost all of the teachers reported having already earned a master's degree (95%) and about half of them had earned it within the last five years. On average, the elementary teachers had been with the district longer (mean = 9.5 years, SD = 8.6) than the middle school teachers (mean = 7.0 years, SD = 6.9). Approximately 50% of the elementary teachers reported being at the same District #2 school for 6.5 years or less and 50% of the middle school teachers had been at the same school for less than 5 years. We lack the data necessary to ascertain whether or not this sample is representative of the District #2 teaching population as a whole.

Summary

Essentially, the 1,536 students associated with the teachers who completed usable questionnaires are reasonably representative of District #2 as a whole. They have a smaller proportion of impoverished students, particularly in grades 6-8, and they have more homogenous test scores. Otherwise, they can be considered a reasonable sample of the District #2 student population.

THE CHALLENGE

Overall, the achievement of District #2 students in the spring of 1999 was strong with a mean performance of 694.84 in mathematics (SD = 66.21) and 690.44 in reading (SD = 55.03). (See Figure 3.) Such strong performance might be used as evidence to support District #2's strategy for instructional improvement, including its system of professional development. Skeptics, however, point to the lower than average poverty rates for an urban district enjoyed by District #2 and ask if the greater proportion of affluent students are the source of their success. Indeed, the relationship between 1998 school performance in reading (percentage of students achieving at or above grade level) and school affluence (percentage of students *not* eligible for free and reduced lunch) throughout New York City's 32 community school districts was quite strong (r = 0.89). (See figure 4.)

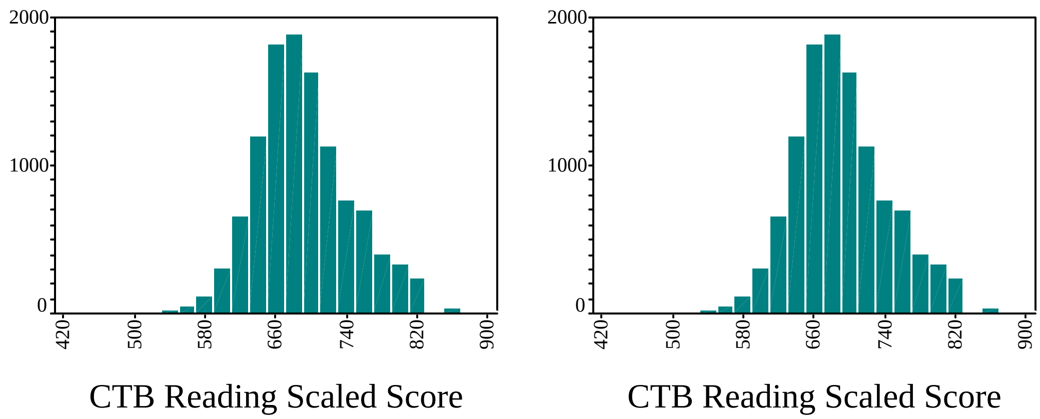

Figure 3: Plots of 1999 CTB scores in mathematics and reading for all District #2 students in grades 3-8

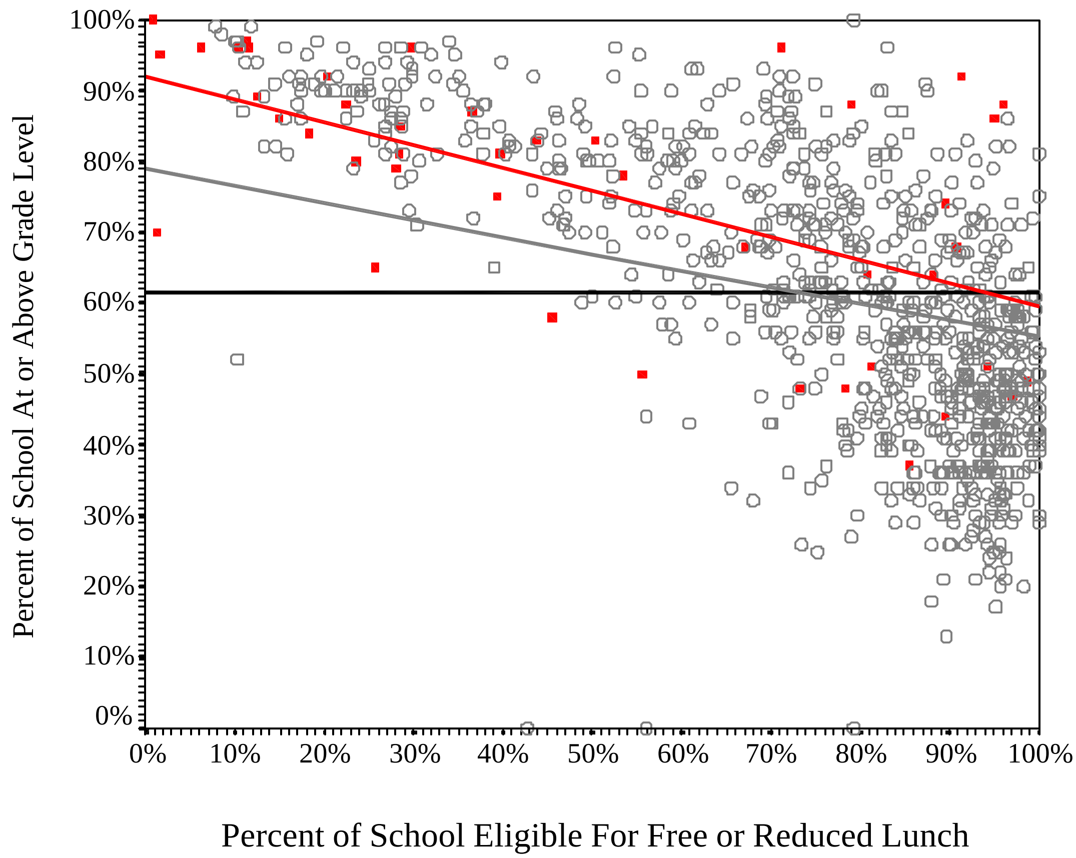

Figure 4: Relationship between school poverty and school performance on the 1998 CTB in reading throughout New York City's community school districts[6]

[6]Data obtained from School Report Card database made available by New York City's Board of Education (www.nycenet.edu) and available on-line at http://207.127.202.63.

District #2 schools (the dark squares in Figure 4) follow the general trend seen in schools throughout New York City. Those schools with lower proportions of students on free and reduced lunch are more likely to have a larger percentage of their students achieving at or above grade level on the CTB in reading.[7] The upper slanted line in Figure #4 represents a simple linear regression for SES fitted to District #2 data. The lower slanted line represents a fitted regression line for the remaining New York City schools and the horizontal line represents the city-wide average percentage of students per school which perform at or above grade level. These data show that some, but not all, of District #2 schools with high levels of students on free and reduced lunch appear to be doing better than expected. Thus, while overall District #2 students appear to be achieving better than average, performance gaps still exist.

If all students are to achieve high standards, then these gaps must close. We review the 1999 CTB data on student achievement in mathematics and reading to determine the size of these gaps and therefore the challenge before District #2's professional development system. Below we describe average achievement by gender, ethnicity, socio-economic status, English proficiency and time in school, as well as the variation in achievement linked to these factors.

Gender

The achievement of District #2's male and female students in grades 3-8 is fairly similar. The differences between their achievement on average are small—2 or 3 points in mathematics and 8 or 9 in reading and they account for less than 1% of the variance on each test. (See table 4).

	Mathematics	Reading
Grades 3-5		
Female	670.27	678.89
Male	673.84	670.16
Grades 6-8		
Female	717.90	711.96
Male	719.56	707.92

Table 4: Mean CTB scores in mathematics and reading by gender of District #2 students in grades 3-8

[7]The relationship between achievement and eligibility for free and reduced price lunch are similar in mathematics and in all three years available (1996, 1997 and 1998) for which data are via the School Report Card database.

Ethnicity

Achievement varies among students of different ethnic backgrounds. (See table 5.) White students score about 40 points higher than Black and Hispanic students in reading, while White and Asians score about 30 points higher than Blacks and Hispanics in mathematics. For students in grades 3-5, ethnicity accounts for 10.4% of the variance in mathematics scores and 10.3% of the variance in reading scores. For students in grades 6-8, ethnicity accounts for 7.4% of the variance in mathematics scores and 6.8% of the variance in reading scores.

	Mathematics	Reading
Grades 3-5		
Asian	681.47	669.72
Hispanic	645.26	656.21
Black	646.60	660.95
White	689.86	695.60
Grades 6-8		
Asian	726.61	707.21
Hispanic	694.31	689.79
Black	700.88	697.58
White	737.43	727.02

Table 5: Mean CTB scores in mathematics and reading by ethnicity of District #2 students in grades 3-8

Socio-Economic Status

Low SES students (those eligible for free or reduced price lunch) tend to score approximately 30 points less on the CTB in both reading and mathematics than high SES students, regardless of grade. (See table 6.) Socio-economic status accounts for 6% of the variance in mathematics scores of 3-5 graders and 4.8% of the variance of 6-8 graders. It accounts for more of the variance in reading— 12.2% of the 3-5 graders scores and 8% of the 6-8 graders. The correlation between SES and mathematics achievement ($r = 0.22$) and between SES and reading ($r = 0.31$) is significant. These correlations are similar to those reported in Ensiminger and Slusarcick (1992) and Entwisle and Alexander (1992).

	Mathematics	Reading
Grades 3-5		
Low SES	659.57	659.64
High SES	689.07	694.53
Grades 6-8		

Low SES	706.66	693.93
High SES	735.27	725.54

Table 6: Mean CTB scores in mathematics and reading by SES of District #2 students in grades 3-8

English Proficiency

Students proficient in English tended to score approximately 50 points higher on the mathematics CTB and 70 points higher on the reading CTB than those who were not. Clearly students who are proficient in English have an easier time reading and therefore performing on the achievement tests. Despite these large differences, English proficiency accounts for little of the variance in elementary students scores—only 1.9% in mathematics and 3.1% in reading. It explained slightly more of the variance in the grade 6-8 scores—5.3% in mathematics and 7.3% in reading. These results are attributable to the large difference in the number of students who were or were not proficient in English. For example, for elementary school students, the number of proficient and not proficient students in mathematics was n = 162 and 5888, respectively, and n = 192 and 5900, respectively, for reading. That is, the overall means for reading and mathematics were almost identical to the means for students proficient in English, producing small explained variance statistics.

	Mathematics	Reading
Grades 3-5		
Not proficient	622.53	605.07
Proficient	673.55	675.76
Grades 6-8		
Not proficient	672.01	635.31
Proficient	723.43	711.06

Table 7: Mean CTB scores in mathematics and reading by English proficiency of District #2 students in grades 3-8

Time in school

District #2's strategy for improving instruction will have little affect on students' achievement unless those students attend school. Moreover, it is presumed that the longer a student has been in District #2, the more likely it is that they will have benefited from the improvement efforts. To test these suppositions, we review the attendance rates and enrollment time of students in grades 3-8 and their relationship to achievement in mathematics and reading.

In general students who have spent more time in District #2 perform better on the CTB in both mathematics and reading. Most enrolled students attend school regularly (mean = 94.2% of days school is in session, SD = 5.96) regardless of gender, socio-economic status or English proficiency. The attendance rate of Asian students is a bit higher than the rest, at approximately 98%, but otherwise there are no differences in attendance rates by ethnicity. Perhaps because so many of the students attend regularly, the correlation between attendance and achievement in mathematics ($r = 0.162$) and reading ($r = 0.080$), although statistically significant, is very small.

Likewise, in the spring of 1999 most students in grades 3-8 had been enrolled in District #2 for at least three years (88.4%), 10.4% had been enrolled for just 2 years and 1.2% had been enrolled for only one year. The mean achievement scores of students in both mathematics and reading were higher for those students who had been enrolled in District #2 longer. (See figure 5.) However, the variance explained by this variable was negligible for both mathematics and reading (2%). As with English proficiency, the number of students who have been enrolled in District #2 for at least three years is so large compared to those who are new to the district, that their average achievement rate is nearly the same as the overall average achievement. Thus, little variance is explained by time in the district alone.

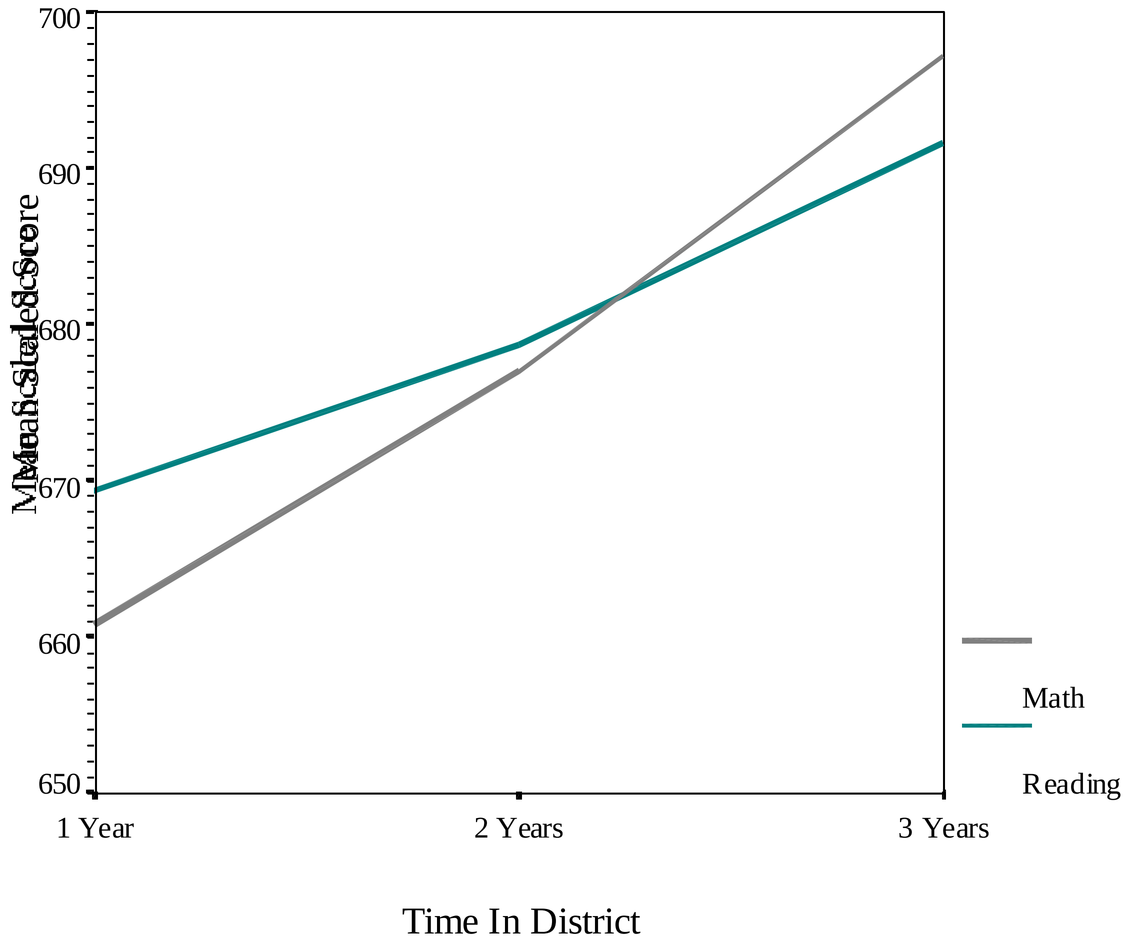

Figure 5: Mean achievement of students compared to time enrolled in District #2 schools

Variation in achievement patterns by classroom

The descriptive statistics reviewed above show that while there is no appreciable difference in the test performance of boys and girls, District #2 does see achievement differences between students of different socio-economic, ethnic and linguistic backgrounds. Differences in achievement between ethnic groups are most pronounced in the early grades, between the poor and affluent in the subject of reading, and between those proficient and not proficient in English in grades 6-8. Attendance appears to improve performance slightly and number or years enrolled in District #2 to improve it moderately.

Inferential analyses conducted with multilevel models on the sample of students for whom we have teacher questionnaire data provide a slightly different picture. (See Appendix C.) These analyses take into account the relationships between socio-economic status, gender, ethnicity, English proficiency, attendance and District #2 enrollment both within and between individual classrooms. The results of the within classroom analyses indicate that on average SES, ethnicity and time enrolled in District #2 were moderate to moderately strong predictors of mathematics achievement, while gender, time enrolled in District #2 and socio-economic status were moderate to strong predictors of reading achievement. (See table 8.)

	Effect on Mathematics	Effect on Reading
SES	Moderately strong	Strong
Ethnicity	Moderate	None
Gender	None	Moderate
English proficiency	—	—
Time enrolled in District #2	Moderately strong	Moderately strong
Attendance	None	None

Table 8: Average within-classroom predictive strength of various student factors

Between classroom analyses showed that the relationship between achievement and attendance did not vary by classroom in either mathematics or reading. The relationship between reading achievement and most of the other factors described here did vary between classrooms. Only the relationship between reading achievement and English proficiency did not. In contrast, the relationship with socio-economic status was the *only* one by which mathematics achievement varied between classrooms. In other words, in some classrooms the relationship between socio-economic status and mathematics achievement is strong, while in others it is weaker. In the next section we examine the extent to which teachers' engagement in professional development can explain the variation in these relationships between classrooms. That is, we ask the question, "Are teachers with strong professional development participation patterns more likely to have closed achievement gaps?"

THE ROLE OF PROFESSIONAL DEVELOPMENT

In this section, we first describe teachers' average participation in professional development activities based on their responses to the questionnaire. Then we discuss findings on the relationship between engagement in these various activities and student achievement patterns in both literacy and mathematics. Finally, since District #2 has made an explicit decision to spend more money on professional development at the expense of reducing class size, we examine the extent to which class size relates to student achievement.

Extent to which District #2 teachers engage in various professional development activities

Teachers were asked to respond twice to six questions about the frequency of their participation in professional development activities in their school—once with respect to mathematics and once with respect to literacy. The activities described were all ones that are part of the regular daily routine in District #2 schools, and not special events, workshops or inservices. Teachers indicated how frequently they engaged in these activities in the last academic year (1998-99): daily, weekly, monthly, 1 or 2 times, never. These data were coded as 5 = daily to 1 = never. (See Table 9.)

	Mathematics		Literacy	
How often in the past year did you...	Mean	Std. Dev.	Mean	Std. Dev.
Have detailed discussions about instructional practice with other teachers in your school? (5a & 5g)	3	1.17	3.81	1.01
Plan lessons with other teachers in your school? (5b & 5h)	2.58	1.25	3.18	1.17
Observe another teacher's practice in your school? (5c & 5i)	2.06	1.13	2.45	1.02
Have another teacher observe your practice and provide feedback? (5d & 5j)	1.65	.93	1.97	1.06
Have a professional developer observe your practice and provide feedback? (5e & 5k)	1.77	.98	2.24	1.13
Talk with your principal about your practice? (5f & 5l)	1.76	.82	2.61	1.03

Table 9: Mean response to items about engagement in the school professional culture (5 = daily; 4 = weekly; 3=monthly; 2=1 or 2 times; 1 = never)

Teachers were also asked to indicate whether or not they participated in a variety of professional development activities in the district as a whole. These activities tended to be in the form of particular roles (e.g. mentor teacher) or special events (e.g. a three week participation in the Professional Development Lab). Teachers were asked to indicate if they had participated in the described activity (yes) or not (no). Five of these activities were divided into two kinds: (1) those District professional development activities in which the teacher acted as a *receiver* of assistance (PcReceive) and (2) those in which the teacher acted as a *provider* of assistance. All five of these activities can be considered forms of "high maintenance" professional development, as they require significant coordination and commitment on the part of District #2 personnel. Three items fell into the *PcReceive* category:

- work with a mentor teacher;
- participate in the professional development lab (PDL) as a visiting teacher; and
- Observe another teacher at work in another school in your district that was not part of a PDL experience.

Two items fell into the *PcProvide* category:

- serve as a mentor for a novice teacher; and
- participate as a PDL resident teacher.

Each teacher was given a score for the number of PcReceive items for which they indicated "yes" and a score of the number of PcProvide items for which they indicated "yes". (See Table 10.)

	Mean	Std. Dev.
PcRecieve	.37	.58
PcProvide	.71	.71

Table 10: Mean response to items about engagement in District #2 professional community

These data indicate that professional development participation rates are higher in literacy than in mathematics. Teachers report on average nearly weekly discussions of instructional practice in literacy and monthly opportunities to plan literacy lessons with other teachers in their school. In mathematics, however, teachers are more likely to engage in the described professional development activities on a monthly or occasional basis. Moreover, only about one third of the teachers had participated as a recipient in any of the district's "high maintenance" professional development in the past year while approximately three quarters had participated at least once as providers.

Influence of professional development on student achievement

The multilevel analyses on the sample of student data for which we also have teacher questionnaire responses indicates that only a few of the professional development activities described above explain any of the between classroom differences in achievement patterns. (See Appendix C for details.)

We are unable to explain any of the classroom variation in mathematics achievement. While classroom achievement in mathematics does vary, none of the classroom level predictors explain that variance. For example, none of the teacher professional development experiences explain it, nor do classroom aggregates of student variables, such as percent of students eligible for free and reduced lunch, explain it. Moreover, while the relationship between mathematics achievement and socio-economic status varies by classroom, engagement in professional development as captured by the questionnaire does not explain that variation.

There is both more classroom variation and more ability to explain that variation in reading achievement. Classroom aggregation of socio-economic status (i.e. the percent of high SES students in a classroom), does predict average student performance in reading. In other words, classrooms with high proportions of high SES students are more likely to have higher classroom averages on the CTB in reading.

Teachers engagement in two of the professional development activities in literacy are also predictive of average classroom achievement. Teachers who discuss literacy instruction more frequently with other teachers are *less* likely to have high classroom averages on the CTB in reading. This relationship disappears, however, when the years of teaching experience of each respondent are taken into account. Thus, it appears that less experienced teachers both tend to have lower achieving students *and* to talk about their literacy instruction more than experienced teachers. Being observed by other teachers and reflecting with them on performance, however, is associated with higher classroom performance. Those teachers who engage in this activity more frequently have higher average classroom scores in reading.

Finally, we reviewed the ability of engagement in professional development to explain why relationships between achievement and other factors, such as socio-economic status, ethnicity, gender, and English proficiency, varies by classroom. As we mentioned earlier, the relationships between reading achievement and socio-economic status, ethnicity, gender, and number of years students were enrolled in District #2 all varied by classroom. However, only one of these relationships could be explained by teacher engagement in professional development — gender and reading achievement. The relationship between gender and achievement was weaker in those classrooms where the teacher discussed their literacy instruction with colleagues frequently. In contrast, socio-economic status was the only relationship with mathematics achievement that varied by classroom and engagement in professional development activities was not able to explain any of that variation.

In summary, engagement in professional development, as measured by this questionnaire and reported by the 62 respondents, does not appear to have significant influence on student achievement in either literacy or mathematics. Only two of the activities (discussing instruction and being observed by others) accounted for any variance in classroom averages for reading achievement. None of the variation in mathematics achievement was explained by engagement in professional development as reported by the teachers. In addition, engagement in professional development did not appear to help to reduce differences often seen in the achievement patterns of students from different socio-economic, ethnic, and linguistic backgrounds. Only the relationship between gender and reading achievement appeared to be mediated by increased engagement in professional development. None of the relationships in mathematics were mediated by engagement in professional development.

Influence of class size on student achievement

Since District #2 made an explicit decision to invest in professional development rather than class size reduction, we decided to investigate the overall effect of class size on student achievement. Class size does vary throughout the district, due in part to naturally occurring enrollment fluctuations, and in part to special funds provided to the district *explicitly* for the purpose of class size reduction. We ran two multilevel models to explore the relationship between class size and student performance. (See

Appendix C for details.) Findings in the literature on class size reduction indicate that this strategy is most effective for students in grades 1-3 (Finn, 1998; Finn & Achilles, 1990; Finn & Voelkl, 1994). In order to compare our results to this literature, our first model included only the students in the 103 third grade classrooms in District #2. The second included the 478 classrooms which contained students in grades 3-8. The results in mathematics and reading from both models are similar. They show that with all other predictors held constant, each additional student in the classroom will *increase* average student performance per classroom by about one point. In short, for all practical purposes, class size does not appear to have any relationship with classroom level student achievement in District #2. In other words, District #2's decision to spend its discretionary funds on professional development rather than class size reduction has had no deleterious effects on student achievement.

CONCLUSION

In this study we have tried to better understand the influence of a variety of factors on student achievement. As expected, we found that student performance in reading and mathematics in District #2 does differ based on students' poverty rates, their ethnicity, and their proficiency in English. We were able to find variance in the strength of several of these relationships among classrooms. However, teachers' reported engagement in professional development was found to make a difference in only one area: the relationship between gender and reading achievement was weakened when teachers discussed their literacy instruction with colleagues frequently.

Why did professional development not show more powerful effects in reducing the achievement gaps? There are many possible explanations, some methodological and some conceptual. Methodologically, the survey sample, although fairly representative with respect to students (i.e., the students taught by the teachers in the sample were very similar to students district-wide), may not have been representative with respect to teachers. With only 12% of the possible teachers represented, we cannot be sure that the sample was not perhaps somehow skewed with respect to potentially important variables such as length of time in the district. The available data did not permit us to ascertain teacher representativeness. In addition, we had no means of checking the reliability of teachers' reported engagement in professional development on the survey. Teachers may have been swayed to answer in ways that were socially desirable and/or may have had difficulty accurately remembering some of their professional development experiences.

Conceptually, the lack of powerful effects from professional development may be explained by how we defined professional learning—as an attribute of the individual teacher rather than as a characteristic of the overall climate of professional culture in a given school. In District #2, the emphasis has been on establishing a school-wide culture of learning for adults as well as for children. Not only are all teachers expected to continue to learn, but they are also expected to assist the learning of other teachers and to assume responsibility for the learning of all the children in their school, not just those in their individual classrooms. If this is indeed the case, one might expect to find the effects of professional development spread out more among children throughout the school as opposed to neatly tagged between one teacher and her students. Despite the fact that our survey tried to measure "participation in a community," we may need to think about additional measurements of the strength of the learning community to which an individual teacher has access.

On the whole this study has provided a good foundation for future studies that will measure the linkages between and among various aspects of the District #2 theory of action—not only professional development—and student outcome measures. Future work will take into consideration overall building climates, as well as the actual practices of teachers in the classroom as additional mediating influences on student achievement.

REFERENCES

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage.

Bryk, A. S., Raudenbush, S. W. & Congdon, R. T. (**1996**). *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Scientific Software Incorporated.

Council of the Great City Schools. (1999). Characteristics of Great City School Districts: Demographic and special needs information about Council districts. Available on-line at http://www.cgcs.org.

CTB. (1991). Comprehensive Test of Basic Skills. Monterey, CA: CTB/McGraw-Hill.

Elmore, R. F. & Burney, D. (December, 1996). Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City. New York, NY: National Commission on Teaching and America's Future.

Elmore, R. F. & Burney, D. (December, 1998). Continuous Improvement in Community District #2, New York City. A report to the Office of Educational Research and Improvement. Pittsburgh, PA: High Performance Learning Communities Project, Learning Research and Development Center, University of Pittsburgh.

Ensminger, M.E., & Slusarcick, A.L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*, 95-113.

Entwisle, D.R., & Alexander, K.L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*(1), 72-84.

Finn, J.D. (1998). Class size and students at risk: What is known? What is next? Paper prepared for the National Institute on the Education of At-Risk Students, Office of Educational Research and Improvement, U.S. Department of Education.

Finn, J.D., & Achilles, C.M. (1990). Answers about questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 557-577. 1990.

Finn, J.D., & Voelkl, K.E. (1994). Class size. In T. Husen & T.N. Postlethwaite (Eds.), *International Encyclopedia of Education*, (2nd edition). Oxford: Pergamon Press.

Gay, L. R. (1996). *Educational research: Competencies for analysis and application, 5th edition.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Grolnick, W.S., Benjet, C., Kurowski, C.O., & Apostoleris, N.H. (1997). Predictors of parent involvement in children's schooling. *Journal of Educational Psychology, 89*(3), 538-548.

Harwell, M.R., & Gatti, G.G. (1999, April). Modeling student achievement in Community School District #2 in New York City. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

High Performance Learning Communities Project. (In progress). HPLC Technical Report I. Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA.

Kennedy, M. (1998). *Form and substance in inservice teacher education* (Research monograph 13). Madison, WI: National Institute for Science Education, University of Wisconsin—Madison.

Lee, V.E., & Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*, 172-192.

Newmann, F.M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco: Jossey Bass.

Raudenbush, S.W., Rowan, B., & Cheong, Y.F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30*(3), 523-553.

Resnick, L. B., & Harwell, M.R. (1998). Professional development and teaching quality in a standards referenced education system. Report to the U. S. Department of Education, Office of Educational Research and Improvement on the UCLA-CRESST project. University of California at Los Angeles, Center for Research on Evaluation, Standards, and Student Testing.

Stein, M.K., D'Amico, L., & Johnstone, B. (April, 1999). District as professional educator: Learning from and in New York City's community school district #2. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Stein, M.K., Harwell, M., & D'Amico, L. (1999). Toward closing the gap in literacy achievement. High Performance Learning Communities Project, Report #001 (4.5.b), Learning Research and Development Center, University of Pittsburgh.

Stein, M. K. & Henningsen, M. (1992, April). Using triangulation in classroom research: An example from small groups in mathematics instruction. Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

APPENDIX A

Teacher Background Survey & Consent Form
High Performance Learning Communities Project
Learning Research and Development Center, University of Pittsburgh

Dear District #2 teacher:

District #2 has entered into a 5 year collaboration with the University of Pittsburgh to improve and study the nature of the district's learning communities. As District #2 moves toward the development of a standards-based learning culture, the University of Pittsburgh researchers are assisting and studying the process at all levels: the district, the schools and the classrooms. As a professional member of the District #2 community, you are being asked to complete the attached questionnaire which is a part of the overall research study.

There are no foreseeable risks associated with this study and the District will benefit from learning more about classroom and school reform. Any information obtained about you from this questionnaire will be utilized for research only and will be kept confidential. Information about you will be coded and your name will be removed from any report. Information which might link you to the research will be kept in locked files for five years and only the research team will have access to it. There is no adverse effect if you choose not to complete this form.

If you have any further questions about this research, you can contact Nancy Israel at the High Performance Learning Communities Project at (412) 624-7452. Any questions about your rights as a research participant will be answered by the Office of the Senior Vice Chancellor for Health Sciences, University of Pittsburgh (412-647-9834). By signing this form, you agree to participate in this research study.

Please sign this consent form and complete the survey attached to it. Place the completed survey and consent form in the enclosed envelope, seal it and return it to the principal of your school. Please keep the copy of the consent form for your own records.

Thank you for participating in this collaborative effort aimed at improving district learning communities.

Sincerely,

Lauren Resnick
Principal Investigator

Participant's Name

Date

1. Name: ____
(Please print)

2. School: PS____________________ **Official Class Number:** ___
(Please print)

3. Education

a. What is the highest degree you have completed thus far?

- ❑ High school diploma
- ❑ Bachelor degree
- ❑ Master degree
- ❑ Doctorate

b. When did you complete this degree? _____________(year)

c. What educational certifications have you acquired thus far?

Please list all certifications below. Include teaching certifications and others, such as administrative ones. Specify grade levels and subject matter where appropriate.	**Year received**

d. How many courses in mathematics, literacy and assessment have you taken for college or university credit? (Include only those courses for which you received credit toward either an undergraduate or graduate degree. Please *exclude* workshops and other professional development activities for which you did *not* receive credit toward a degree.)

Area	**Total number of courses taken**	**Number of these courses taken in the last five years**
a. Mathematics		
b. Mathematics curriculum & methods		
c. Writing		
d. Literature		

e. Literacy curriculum and methods		
f. Assessment and/or testing		

4. Teaching experience

a. How many years have you been teaching in a school that serves K-12 students? ___________

b. How many years have you been teaching in your current school district? ___________

c. How many years have you been teaching at your current school? ___________

d. What grade(s) did you teach this past academic year? ______________________________

e. How many years have you been teaching in your current grade(s)? _________

f. Are you currently teaching a grade level(s) for which you are certified? ❑ Yes ❑ No

g. Are you a subject matter specialist? ❑ Yes ❑ No

If yes, what subject(s) did you teach this year? ______________________________

If yes, how many years have you been teaching your current subject(s)? _________

h. Are you currently teaching a subject(s) for which you are certified? ❑ Yes ❑ No

5. Professional community in your school

	On average, how often in the past academic year did you...	**Daily**	**Weekly**	**Monthly**	**1 or 2 times**	**Never**
a.	Have detailed discussions about instructional practice in **mathematics** with other teachers in your school?	❑	❑	❑	❑	❑
b.	Plan **mathematics** lessons with other teachers at your school?	❑	❑	❑	❑	❑
c.	Observe another teacher's **mathematics** practice in your school?	❑	❑	❑	❑	❑
d.	Have another teacher observe your **mathematics** practice and provide feedback?	❑	❑	❑	❑	❑
e.	Have a professional developer observe your **mathematics** practice and provide feedback?	❑	❑	❑	❑	❑
f.	Talk with your principal about your **mathematics** practice?	❑	❑	❑	❑	❑
g.	Have detailed discussions about instructional practice in **literacy** with other teachers in your school?	❑	❑	❑	❑	❑
h.	Plan **literacy** lessons with other teachers at your school?	❑	❑	❑	❑	❑

i.	Observe another teacher's **literacy** practice in your school?	❑	❑	❑	❑	❑
j.	Have another teacher observe your **literacy** practice and provide feedback?	❑	❑	❑	❑	❑
k.	Have a professional developer observe your **literacy** practice and provide feedback?	❑	❑	❑	❑	❑
l.	Talk with your principal about your **literacy** practice?	❑	❑	❑	❑	❑

6. Professional community in your district

In the past academic year, did you...

a.	Work with a mentor teacher?	❑ Yes	❑ No
b.	Serve as a mentor for a novice teacher?	❑ Yes	❑ No
c.	Participate in the professional development lab (PDL) as a visiting teacher?	❑ Yes	❑ No
d.	Participate in PDL as a resident teacher?	❑ Yes	❑ No
e.	Observe another teacher at work in another school in your district that was not part of a PDL experience?	❑ Yes	❑ No
f.	Participate in the Standards Network?	❑ Yes	❑ No
g.	Serve as a mathematics lead teacher?	❑ Yes	❑ No

7. Professional community beyond your district

In the last *two* years, did you...		**...in mathematics**		**...in literacy**	
a.	Attend a workshop or conference outside your district?	❑ Yes	❑ No	❑ Yes	❑ No
b.	Make a presentation at a conference outside your district?	❑ Yes	❑ No	❑ Yes	❑ No
c.	Lead or facilitate a workshop or inservice session on improving teaching practice outside your district?	❑ Yes	❑ No	❑ Yes	❑ No

8. Background information

a. When did you graduate high school? ___________(year)

b. What is your ethnic background? (Please check all that apply.)

❑ Black

❑ White

- ❑ Asian, Asian-American, or Pacific Islander
- ❑ Hispanic, Puerto Rican, Latino or Chicano
- ❑ Native American or Alaskan Native
- ❑ Other (please specify): ______________________

c. What is your gender?

- ❑ Male
- ❑ Female

APPENDIX B
DATA PROBLEMS AND ISSUES

Missing data

11% of the students were missing achievement scores in mathematics and 14% were missing them in reading. 55.7% of those students with missing test scores in mathematics and 79% of those missing scores in reading were also English Language Learners. These students were most likely exempt from testing due to their lack of English proficiency. Asian students were more likely to be missing test data than other students. In fact, they accounted for 62% of the missing data in mathematics and 69% of the missing data in reading, and overall 26.5% of the Asian students were missing test scores in mathematics and 18.5% were missing them in reading. 87% of those Asian students with missing achievement data in reading and 55% in mathematics are considered English Language Learners. This suggests that reading scores are higher than they would be if English Language Learners were not exempted from these tests.

In addition, the teachers' questionnaires were not always complete. Two teachers did not respond to one of the twelve questions (5a-5l) in the section on their engagement in "Professional community of your school"; six teachers did not respond to one of the seven questions (6a-g) on their engagement in the "Professional community of District #2"; and quite a few did not respond to the three questions (7a-7c) about engagement in the "Professional community beyond District #2". No teacher missed answering more than one of the questions in the first 19 of these questions (5a-5l and 6a-6g). Items 7a-7c were not used in analysis and the missing values for items 5a-5l and 6a-6g were imputed from the teachers' modal response. Where possible, we ran analyses with and without the imputed values, and found no difference in the pattern of findings. As a result, we report results using the imputed responses for teachers.

Lack of variation in some classrooms

Several of the 62 classrooms for which both teacher and student data was available show no variation on key variables. For example, 13 classrooms showed no variation in the socio-economic status of their students—they were either all eligible for free or reduced lunch, or they were all not eligible. In four classrooms, all the students had been enrolled in the district for the same number of years. Multilevel models are designed to work with quantitative variables and hence ethnicity was recoded from four categories into two. We chose to collapse the categories into white and non-white because early analyses showed that the achievement of white students differed significantly from the rest of the students. Once this recategorization was made, 13 of the classrooms showed no variation in ethnicity, consisting of either all white students or all non-white students. Finally, just over half of the classrooms (33) showed no variation in English proficiency. Because multilevel analyses are designed to explain variation at different levels of the system, lack of variation within particular classrooms on any given variable excluded that classroom and all the students within it from the analysis. Since English proficiency affected more than half of the classrooms, it was dropped from all analyses as a student-level predictor. Even with this provision, most of the multilevel analyses could only be run on a portion of the data in the sample. Actual numbers of classrooms and students used in the various analyses differed, but in general included 30 classrooms and approximately 600 students.

The loss of these classrooms and students from the HLM analyses has three consequences. First, it introduces bias into our results. Of the 13 classrooms with no variation in SES, 9 were composed entirely of low SES students. Likewise, those with no variation in ethnicity tended to be nonwhite, and those with no variation in the number of years students had been enrolled in District #2 tended to have been in the district at least three years. Second, our ability to generalize from our findings is weakened, since they exclude homogenous classrooms. Finally, the precision with which classroom-level effects can be estimated was reduced. These problems are somewhat endemic of school-based research and are reported regularly in educational literature (e.g. Newmann & Associates, 1996).

Connecting students in grades 6-8 with their middle school teacher

Finally, we are still encountering some data coordination problems at the middle school level. Student and teacher data is matched based upon the teacher ID number associated with each student. For elementary students, this teacher is the classroom teacher with which the student spends the majority of their day and who provides instruction in both mathematics and literacy. For students in grades 6-8, it is less clear what role the teacher listed plays. It is possible that the teacher listed is a homeroom teacher of some sort and may not be the one who provides instruction in either literacy or mathematics to the students assigned to them.

APPENDIX C
STATISTICAL ANALYSES

We used regression models/hierarchical linear models (HLM) (Bryk & Raudenbush, 1992) to provide evidence about our original research questions:

(1) What are the effects of predictors of student achievement that District #2 *cannot* control (SES, Ethnicity, Gender, English Proficiency, Years a Student Has Been In District #2, Attendance) within classrooms?

(2) Are the effects of the predictors in (1) weakened by teachers' reported engagement in professional development?

(3) Is student achievement affected by class size?

HLM Analyses

To address the research questions a series of multilevel analyses (students nested within teachers/classrooms) were performed using the HLM computer program (Bryk, Raudenbush, & Congdon, 1996). Under the assumption that i=1,2,.., n_j students (level 1) are nested within j=1,2,…,J classrooms (level 2), and that District #2 students and classrooms are representative of an identifiable population, we fitted regression models to mathematics scores and, separately, to reading scores. The average within-classroom sample size in these analyses was 22 and ranged from 8 to 34. We examined residuals for the fitted models and found no serious evidence of nonnormality and heteroscedasticity, although we did find evidence of nonlinearity associated with the Years a Student Has Been In District #2 or Time In District (TID) variable that appeared to be quadratic in nature. In these cases, we used two strategies (a) We created a new variable TID^2 that was included in our analyses (b) We performed a log-transformation for TID to try to reduce the quadratic effect and used this variable in our analyses. The results of our analyses using TID^2 and log(TID) were similar to those based on TID, and only the latter are reported. Plots of the residuals against products of the predictor variables suggested that interaction-type predictors were not needed.

We used a significance level of .05 for all statistical tests.

We began by fitting an unconditional model that allowed us to explore the total variation between classroom means. The initial level 1 and level 2 models were:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad (1a)$$

$$\beta_{0j} = \gamma_{00} + u_{00}$$

In equation (1a), Y_{ij} is a test score for student i in classroom j (i =1,2,..,n_{jk}; j=1,2,..,J_k), β_{0j} is the mean achievement of classroom j, r_{ij} is an error term that is assumed to follow a normal distribution $N(0, \sigma^2)$ and constant variance, and u_{00} represents the difference between each student's test score and the classroom mean and is a random effect.

Next, we modeled variation in classroom mathematics and readings means using classroom-level predictors. The level 1 model was the same as (1a) but the level 2 model was

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad (1b)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{01} W_{2j} + \ldots + \gamma_{0j} W_{tj} + u_{0j}$$

where β_{0j} is the mathematics or reading mean of the jth classroom, γ_{00} is the intercept, γ_{01} is a slope capturing the relationship between a classroom level predictor such as %High SES students and classroom achievement, and u_{0j} is a random error that is assumed to follow a normal distribution $N(0, \tau_{00})$. Although we would have preferred to perform these analyses separately for elementary and middle school teachers, the small number of middle school teachers (15) made this strategy unfeasible. Instead, we included Grade as a level 2 predictor.

Finally, we examined the effects of predictors of student achievement by fitting a level 1 regression model to the data for each classroom:

$$Y_{ijk} = \beta_{0j} + \beta_{1j}\ \text{SES} + \beta_{2j}\ \text{Ethnicity} + \beta_{3j}\ \text{TID} + \beta_{4j}\ \text{Gender} + \beta_{5j}\ \text{Attendance} + r_{ij} \qquad (1c)$$

where β_{1j} is the slope capturing the relationship between SES and achievement. All level 1 predictors were centered about their group means. The results from fitting models (1b) and (1c) to the data provided evidence about our first research question. We then attempted to explain variation among classroom intercepts and slopes using a level 2 model of the form:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{01} W_{2j} + \ldots + \gamma_{0j} W_{tj} + u_{0j} \qquad (1d)$$

$$\beta_{tj} = \gamma_{10} + \gamma_{11} W_{1j} + \gamma_{21} W_{2j} + \ldots + \gamma_{tj} W_{tj} + u_{1j}\ ,\ t = 1,2,..,\ T \text{ level 2 predictors}$$

where g_{00} is the mean achievement across classrooms, g_{01} is slope relating the level 2 predictor W_{1j} to level 1 classroom mean b_{0j}, u_{0j} is the effect of classroom j on mean achievement conditioning on level 2 predictor W_{tj}, g_{10} is the intercept for level 2 regressions, g_{tj} is a slope relating the level 2 predictor W_{tj} to level 1 slope b_{pj} , and u_{1j} is the effect of a classroom on the level 1 slope conditioning on W_{tj}. Level 2 predictors included %high SES, %White, %English Proficient, and Grade level, as well as variables based on the survey that teachers responded to. All level 2 predictors were centered about their grand mean.

Essentially, the model for b_{0j} in (1b) and (1d) attempt to predictor variation in classroom means as a function of classroom level predictors like %High SES and teachers reported professional development. The model for b_{tj} in (1d) attempts to predict variation in level 1 slopes relating level 1 predictors like SES and achievement as a function of classroom level predictors. The results from fitting model (1d) to the data provided evidence about our second and third research questions.

Mathematics Results

The results of fitting model (1a) to the mathematics data are summarized in the upper portion of Table 11. The resulting chi-square test for mathematics (c^2 = 2,614.3, p < .00) indicates that there was significant variation among the mathematics means, with 64% of the variation among the means attributable to between-classroom variation. These results provide evidence that student achievement in mathematics varies across classroom.

Next, we fitted model (1b) to the classroom mathematics means with the predictors %High SES, %White, %English Proficient, Grade, PcProvide, PcReceive, and items 5a-5l from the teachers survey. The results are reported in Table 11 and indicate that only the predictor Grade was significant. Although the residual variance between classroom is smaller than that associated with model (1a) (826.53 compared to 1214.48), the size of a few of the nonsignificant slopes suggests that the power associated with these tests is quite low.

Next we fitted model (1c) to the data. Initial analyses indicated that Attendance and Gender were not significant predictors of mathematics scores, and model (1c) was refitted to the data after deleting these predictors. The results of these analyses are based on 30 classrooms and are summarized in the middle portion of Table 11. The median R among the regression models fitted within classrooms was .54 and the median R^2 was .29.

The slopes associated with SES, Ethnicity, and TID were all statistically significant predictors of mathematics performance. For example, the slope for the within-classroom regression of mathematics on SES, with the effects of the other predictors held constant, was 7.69, meaning that, on average, each high SES students tend to score about 8 points higher than those in the low SES group. Similarly, the significant slope of 7.87 for Ethnicity means that, with the other predictors held constant, White students tended to score about 8 points higher than Nonwhite students. The largest estimated slope was for TID (15.27), indicating that this is a powerful predictor of mathematics achievement. However, the introduction of these student-level predictors increased the explained variation by only 3.5% compared to the unconditional model in (1). Moreover, the reliability of the estimated slopes was quite low, for example, .15 for SES. However, low reliability for estimated slopes are common in HLM analyses (Bryk & Raudenbush, 1992).

The results for model (1c) in Table 11 also indicate that there was significant variation among the Math/SES slopes (c^2 = 42.93, p = .046), but not the Math/Ethnicity or Math/TID slopes. To study the between-classroom variation among the Math/SES slopes, we fitted model (1d) to the mathematics data with the level 2 predictors classroom %High SES, %White, %English Proficient, Grade, PcReceive, and PcProvide. We then refitted the model to the mathematics data after dropping predictors that showed no evidence of being able to predict variation in the level 1 relationships, and added items 5a-5f as individual predictors. The results for the refitted model (1d) are reported in Table 11.

Grade and %High SES in a classroom were significant predictors of intercepts (classroom mathematics means). The only significant level 2 predictor of level 1 slopes was item 5a (Have detailed discussions about instructional practice in mathematics with other teachers in your school?). The estimated slope for 5a (–6.88) indicates that teachers who reported having detailed discussions about instructional practice in mathematics with other teachers more frequently tended to be associated with classrooms in which the Math/SES slope was weaker. Put another way, the result of teachers engaging in this practice more frequently is a weakening of the grip of SES on mathematics achievement within classrooms.

Table 11

Results For HLM Analysis Of CTB Mathematics Scores

Mathematics Model (1a)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For CLASSROOM MEANS, B0				
INTRCPT2, G00	678.984138	6.060956	112.026	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	47.07290	2215.85823	61	2614.33152	0.000
level-1,	R	34.84945	1214.48436			

Mathematics Model (1b)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	678.922789	3.783834	179.427	0.000
%WHITE, G01	-0.164316	36.520980	-0.004	0.996
GRADE, G02	22.804114	2.938682	7.760	0.000
5A, G03	-0.737961	5.570097	-0.132	0.896
5B, G04	5.512010	6.009691	0.917	0.364
5C, G05	-7.033560	5.492093	-1.281	0.206
5D, G06	-3.642465	6.488816	-0.561	0.577
5E, G07	-5.327072	5.128052	-1.039	0.304
5F, G08	-7.823523	6.159491	-1.270	0.210
%HIGH SES, G09	53.733394	31.306640	1.716	0.092
%ENGPROF, G010	-45.063847	30.398588	-1.482	0.144

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	28.74948	826.53236	51	747.24068	0.000
level-1,	R	34.85845	1215.11148			

Mathematics Model (1c)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	678.979799	6.061532	112.015	0.000
For SES slope, B1				
INTRCPT2, G10	7.697969	2.945394	2.614	0.009
For ETHNICITY slope, B2				
INTRCPT2, G20	7.875013	2.599247	3.030	0.003
For TID slope, B3				
INTRCPT2, G30	15.277287	3.925224	3.892	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	47.10245	2218.64108	29	1344.53073	0.000
SES slope,	U1	8.14074	66.27172	29	42.93701	0.046
ETHNICITY slope,	U2	1.25319	1.57048	29	32.23878	0.309
TID slope,	U3	0.88569	0.78445	29	27.83634	>.500
level-1,	R	34.22332	1171.23563			

Mathematics model(1d)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	678.957613	3.837102	176.945	0.000
GRADE, G01	22.401791	2.925286	7.658	0.000
5A, G02	-3.242360	5.363114	-0.605	0.548
5B, G03	5.236352	6.105981	0.858	0.395
5C, G04	-7.333121	5.704766	-1.285	0.205
5D, G05	-7.726250	6.090243	-1.269	0.211
%HIGH SES, G06	53.363041	12.613592	4.231	0.000
ENGPROF, G07	-43.807756	30.880012	-1.419	0.162
PCRECEIVE, G08	3.261483	6.346245	0.514	0.609
PCPROVIDE, G09	-3.224704	7.338517	-0.439	0.662
For SES slope, B1				
INTRCPT2, G10	9.816033	3.734388	2.629	0.012
GRADE, G11	1.144182	2.187308	0.523	0.603
5A, G12	-6.887267	3.571859	-1.928	0.050
5B, G13	7.219678	5.044912	1.431	0.158
5C, G14	-5.295861	5.463582	-0.969	0.337
5D, G15	3.594222	4.892389	0.735	0.466

```
  %HIGH SES, G16     28.186454      15.678869        1.798     0.078
    ENGPROF, G17   -134.761768      89.059975       -1.513     0.136
  PCRECEIVE, G18      6.995077       5.030018        1.391     0.170
  PCPROVIDE, G19      8.117600       4.991986        1.626     0.110
For ETHNICITY slope, B2
   INTRCPT2, G20      9.977147       4.423901        2.255     0.028
      GRADE, G21     -0.478233       2.278518       -0.210     0.835
         5A, G22     -1.436971       3.584137       -0.401     0.690
         5B, G23     -6.130849       5.400254       -1.135     0.262
         5C, G24      4.987202       4.986903        1.000     0.322
         5D, G25      5.607685       4.665063        1.202     0.235
  %HIGH SES, G26     -3.239829      14.030563       -0.231     0.818
    ENGPROF, G27    -27.801959      86.086324       -0.323     0.748
  PCRECEIVE, G28      3.070890       5.225426        0.588     0.559
  PCPROVIDE, G29      0.354303       5.116938        0.069     0.945
For      TID slope, B3
   INTRCPT2, G30     17.738122       4.680655        3.790     0.001
      GRADE, G31      1.376137       3.897878        0.353     0.725
         5A, G32     -9.389837       5.847281       -1.606     0.114
         5B, G33     -5.614007       7.763638       -0.723     0.473
         5C, G34     11.286958       7.743462        1.458     0.151
         5D, G35     -1.920341       7.849062       -0.245     0.808
  %HIGH SES, G36    -13.133690      19.189643       -0.684     0.497
    ENGPROF, G37     24.532499      86.079413        0.285     0.777
  PCRECEIVE, G38     10.637630       8.320815        1.278     0.207
  PCPROVIDE, G39      1.614742       7.886164        0.205     0.839

Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square   P-value
                        Deviation     Component
-----------------------------------------------------------------------------
INTRCPT1,       U0       29.21685     853.62461    20     425.50978     0.000
     SES slope, U1        4.72651      22.33987    20      31.71373     0.046
ETHNICITY slope, U2       6.48327      42.03282    20      27.40477     0.124
     TID slope, U3        6.72974      45.28945    20      18.91956     >.500
 level-1,       R        34.28529    1175.48099
```

Reading Results

The results of fitting model (1a) to the reading data are summarized in the upper portion of Table 12. The resulting c^2 statistic of 1,757.01 ($p < .00$) tells us that reading means varied significantly across classrooms. We also computed the proportion of the variance in the means that was between-classrooms, .55 or 55%. This value tells us that there is substantial variation among the means that could potentially be explained by level 2 predictors.

We next fitted model (1b) to the reading data for each classroom. Several of the level 2 predictors showed no evidence of being able to predict level 1 slopes and were dropped from the model. The results from the refitted model (1b) are reported in Table 12. In addition to %High SES and Grade, items 5g (Have detailed discussions about instructional practice in literacy with other teachers in your school?) and 5j (Have another teacher observe your literacy practice and provide feedback?) were significant predictors of classroom reading means.

The estimated slope for 5j (11.42) indicates that teachers who have other teachers observe their literacy practice and provide feedback more frequently tend to be associated with higher achieving classrooms.

The estimated slope for item 5g (-10.18) was puzzling because it suggests that teachers who have more detailed discussions tend to be associated with lower performing classrooms. We hypothesized that this finding might be attributable to variation in teaching experience, with the students of inexperienced teachers less likely to profit from this activity. To examine this empirically, we refitted model (1b) after including a predictor reflecting teaching experience (item 4b from the teachers survey). The results supported our hypothesis because the slope associated with 5g was no longer statistically significant; the other results were approximately the same (Adding the teaching experience predictor to the analyses for mathematics described above did not change the pattern of findings).

Next, we fitted model (1c) to the reading data within classrooms. The median coefficient of determination (R) among the regression models fitted within classrooms was .49 and the median R^2 was .24. The results of these analyses are summarized in the middle portion of Table 12. SES, Gender, Ethnicity, and TID were each statistically significant predictors of reading within classrooms. Attendance was not a significant predictor of classroom relationships but we retained it because deleting it noticeably lowered the estimated slopes for SES, Ethnicity, and TID (These smaller slopes were still statistically significant, however). For example, the slope for the regression of reading on SES, with the effects of the other predictors held constant, was 12. 70, meaning that, with the other predictors held constant, high SES students score about 13 points higher on this test than low SES students. Similarly, the slope for TID was 10.73, meaning that, with the other predictors held constant, each additional year in District #2 schools produces a gain of about 11 points on this test. The slope associated with the Ethnicity predictor indicates that on average Whites outperformed NonWhites by about 5 points. Similarly, the slope associated with Gender was statistically significant and indicates that females on average outperformed males by slightly more than 6 points. There was also significant variation among all of the classroom slopes except those for reading/Attendance (The p-value for the test of variation among the reading/TID slopes was .052 and we treated this as evidence of significant variation among these slopes).

Collectively, the results from model (1c) tell us that student-level SES, Ethnicity, TID, and Gender all have a significant effect on reading scores, and that this effect varies across classrooms. We attempted to explain variation among the classroom intercepts and slopes by fitting model (1d) with level 2 predictors %High SES, Grade, %White, %English Proficient, PcReceive, PcProvide, and items 5g-5l from the teachers survey.

A number of the level 2 predictors showed no evidence of being able to predict variation among level 1 slopes and were dropped from the model. Results for the reduced model (1d) are reported in Table 12 (results with and without teaching experience were similar). For the intercept model, Grade and %High SES were again significant predictors of classroom reading means, along with item 5j (Have another teacher observe your literacy practice and provide feedback?). The estimated slope for this predictor of 12.75 means that teachers who engage in this practice more frequently tend to be associated with higher performing classrooms.

For the level 1 slopes, item 5g (Have detailed discussions about instructional practice in literacy with other teachers in your school?) was the only significant predictor from the teachers survey. The estimated slope for 5g (-7.84) tells us that teachers who reported having these detailed discussions more tend to be associated with classrooms with smaller reading/Gender slopes. The other significant level 2 predictor was %English Proficient for reading/Ethnicity slopes. The huge estimated slope of 192.11 means that as students within classrooms become proficient in English the grip of Ethnicity on reading performance strengthens.

The introduction of level 2 predictors also significantly reduced the variance associated with the Reading/Gender slopes. The unconditional variance associate with Gender in model (1c) was 46.84, whereas the conditional variance of Gender after introducing the level 2 predictors in model (1d) was 37.36, a reduction of (46.84-37.36)/46.84 = 20%. Thus, a sizeable portion of the variation in this relationship can be accounted for by teachers reported engagement in these professional development activities.

On the whole, the predictors in model (1c) affected student achievement in reading, with the reading/Gender relationships being predicted by teachers reported engagement in detailed discussions about instructional practice in literacy with other teachers. However, the statistically significant variance components in model (1d) for Gender, Ethnicity, and TID indicate that there is still variance among these slopes to be explained.

Table 12
Results For HLM Analysis Of CTB Reading Scores

Reading Model (1a)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For CLASSROOM MEANS, B0				
INTRCPT2, G00	679.056783	4.703979	144.358	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	36.18119	1309.07880	61	1757.01017	0.000
level-1,	R	32.38429	1048.74204			

Reading Model (1b)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	679.013748	2.706400	250.892	0.000
GRADE, G01	13.889633	1.680130	8.267	0.000
5G, G02	-10.182263	3.352435	-3.037	0.004
5J, G03	11.428636	3.524028	3.243	0.002
5K, G04	3.442595	2.971343	1.159	0.252
%HIGH SES, G05	70.118417	8.534826	8.216	0.000
PCRECEIVE, G06	5.754933	4.229459	1.361	0.179
PCPROVIDE, G07	-6.645972	5.415298	-1.227	0.225

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	19.90510	396.21310	54	482.67618	0.000
level-1,	R	32.38997	1049.11013			

Reading Model (1c)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	679.065427	4.698656	144.523	0.000
For ATTENDANCE slope, B1				
INTRCPT2, G10	0.226386	0.191068	1.185	0.236
For GENDER slope, B2				
INTRCPT2, G20	-6.267913	1.924844	-3.256	0.002
For SES slope, B3				
INTRCPT2, G30	12.703846	2.736711	4.642	0.000
For ETHNICITY slope, B4				
INTRCPT2, G40	5.307353	2.640444	2.010	0.044
For TID slope, B5				
INTRCPT2, G50	10.735700	4.127722	2.601	0.010

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	36.20385	1310.71898	27	563.99252	0.000
ATTENDANCE slope,	U1	0.39734	0.15788	27	25.56035	>.500
GENDER slope,	U2	6.84418	46.84286	27	49.76041	0.005

```
      SES slope, U3         7.69681      59.24085    27       41.02597    0.041
ETHNICITY slope, U4         7.60074      57.77126    27       45.53354    0.014
     TID slope, U5         7.58414      57.51915    27       39.92135    0.052
 level-1,       R         31.08933     966.54670
```

Reading Model (1d)

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	679.164404	2.779510	244.347	0.000
GRADE, G01	13.630251	1.707073	7.985	0.000
5G, G02	-5.717731	4.078141	-1.402	0.167
5H, G03	-4.806944	3.547437	-1.355	0.181
5J, G04	12.755057	3.385114	3.768	0.001
%HIGH SES, G05	69.159124	9.435427	7.330	0.000
ENGPROF, G06	-8.683160	29.604630	-0.293	0.770
For GENDER slope, B1				
INTRCPT2, G10	-5.927174	2.012075	-2.946	0.005
GRADE , G11	1.514231	1.168778	1.296	0.201
5G, G12	-7.848580	2.735535	-2.869	0.006
5H, G13	4.657514	2.474793	1.882	0.065
5J, G14	1.260424	2.384245	0.529	0.599
%HIGH SES, G15	-6.591446	6.694901	-0.985	0.330
ENGPROF, G16	-27.833869	32.167840	-0.865	0.391
For SES slope, B2				
INTRCPT2, G20	13.054464	3.319279	3.933	0.000
GRADE, G21	-0.695852	1.653851	-0.421	0.675
5G, G22	-0.345134	4.126655	-0.084	0.934
5H, G23	-0.730640	3.248692	-0.225	0.823
5J, G24	-2.652110	3.346857	-0.792	0.432
%HIGH SES, G25	26.617589	12.807626	2.078	0.042
ENGPROF, G26	-52.290713	70.322576	-0.744	0.460
For ETHNICITY slope, B3				
INTRCPT2, G30	0.396928	3.961790	0.100	0.921
GRADE, G31	-0.266549	1.685864	-0.158	0.875
5G, G32	-1.414917	4.149533	-0.341	0.734
5H, G33	-1.204034	3.558285	-0.338	0.736
5J, G34	-1.532347	3.509057	-0.437	0.664
%HIGH SES, G35	-9.867138	13.019838	-0.758	0.452
ENGPROF, G36	192.113493	73.555599	2.612	0.012
For TID slope, B4				
INTRCPT2, G40	13.407015	4.704653	2.850	0.007
GRADE, G41	0.221164	2.963696	0.075	0.941
5G, G42	-3.747643	6.127348	-0.612	0.543
5H, G43	-2.768607	5.127370	-0.540	0.591
5J, G44	4.999805	6.242594	0.801	0.427
%HIGH SES, G45	8.516962	18.997742	0.448	0.655
ENGPROF, G46	-77.112832	83.613875	-0.922	0.361

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value

INTRCPT1,	U0	20.47789	419.34387	21	286.71570	0.000
GENDER slope,	U1	6.11247	37.36227	21	41.69862	0.005
SES slope,	U2	7.54027	56.85563	21	30.58146	0.081
ETHNICITY slope,	U3	8.19147	67.10017	21	41.24919	0.005
TID slope,	U4	10.26546	105.37977	21	38.30913	0.012
level-1,	R	31.04718	963.92757			

Analysis of the Effects Of Class Size On Student Achievement

To examine the effect of class size on student achievement we performed two sets of HLM analyses. In one analysis only third grade students in the district were used, consistent with some of the findings in this literature that class size reductions are most beneficial for K-3 students. Our second analysis used all of the 3rd-8th grade students in the district and included Grade as a predictor variable. Both analyses also used %High SES, %White, and %English Proficient as predictors, along with Class Size.

The average number of students in for the District #2 classrooms available to us was 26. We examined the effect of class size for an initial sample of 2,425 third-graders by fitting model (1b) with level 2 predictors to the mathematics and reading data separately. There were no level 1 predictors; we simply wanted to learn if class size was a statistically significant predictor of classroom mathematics and reading means. The results of these analyses are reported in Table 13.

Classroom data for 103 third grade teachers were analyzed. As seen in Table 13, Class Size was not a significant predictor of classroom performance, although other level 2 predictors were. Across grades 3-8, data for 478 teachers were available. As reported in Table 13, Class Size was a significant predictor of mathematics and reading means, although not in the expected way. For mathematics, the estimated slope of 1.12 tells us that with the other predictors held constant, increasing class size by one student will, on average, increase mathematics achievement by about one point, a negligible effect. The estimated slope for reading (1.16) is interpreted in the same way. These results support District #2's decision to not invest its resources in reducing class sizes.

Table 13

Results of Class Size Analyses

Third Grade Students: Mathematics

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	629.930395	1.881022	334.887	0.000
%WHITE, G01	-54.141740	7.390549	-7.326	0.000
%ENGPROF, G02	32.617642	13.205602	2.470	0.014
%HIGH SES, G03	25.155339	5.545321	4.536	0.000
CLASSIZE, G04	0.053286	0.355934	0.150	0.881

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	16.64838	277.16858	93	478.73455	0.000
level-1,	R	35.69181	1273.90560			

Third Grade Students: Reading

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	645.664505	1.534370	420.801	0.000
%WHITE, G01	-16.317332	6.020536	-2.710	0.007
%ENGPROF, G02	20.406238	10.862657	1.879	0.060
%HIGH SES, G03	39.359024	4.511479	8.724	0.000
CLASSIZE, G04	0.076555	0.292560	0.262	0.794

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	13.15547	173.06640	93	387.12844	0.000
level-1,	R	32.53366	1058.4388			

Grades 3-8: Mathematics

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	684.586736	1.609867	425.244	0.000
%WHITE, G01	-55.417917	6.311413	-8.781	0.000
GRADE, G02	19.800177	1.004470	19.712	0.000
%ENGPROF, G03	73.362077	8.706457	8.426	0.000
%HIGH SES, G04	28.567808	5.334608	5.355	0.000
CLASSIZE, G05	1.121817	0.268467	4.179	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	34.21208	1170.46621	477	10293.71979	0.000
level-1,	R	35.93373	1291.23259			

Grades 3-8: Reading

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	680.543434	1.335013	509.765	0.000
%WHITE, G01	-23.454722	5.235822	-4.480	0.000
GRADE, G02	13.202628	0.832914	15.851	0.000
%ENGPROF, G03	67.783668	7.294942	9.292	0.000
%HIGH SES, G04	44.193832	4.418340	10.002	0.000
CLASSIZE, G05	1.161637	0.223262	5.203	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	28.20720	795.64588	477	8804.94870	0.000
level-1,	R	32.28986	1042.63491			