PAHSCI Coaching and Student Outcomes:
Evidence from Philadelphia

Elliott Medrich, Ph.D.
MPR Associates, Inc.

Summary
A study of student performance on benchmark tests at four Philadelphia high schools participating in the Pennsylvania High School Coaching Initiative (PAHSCI) shows some positive results of one-on-one teacher coaching. At two of the four high schools, the students of one-on-one coached teachers had better benchmark test results than the students of the uncoached teachers in control schools. While the magnitude of the differences were not statistically significant, for reasons discussed, the data suggest that the students of teachers who work with coaches, one-on-one, on a regular basis, and who also receive instruction in evidence based literacy strategies, perform better on the Philadelphia Benchmark tests than students of teachers who are uncoached. While results varied by school, and by subject, the data suggest some real advantages for coached teachers and their students.

---

The Pennsylvania High School Coaching Initiative (PAHSCI) is a high school reform program that places one literacy and one math coach per 600 students in high-needs high schools in 15 Pennsylvania school districts.¹

Twenty-four high schools across the state participate in PAHSCI, including four comprehensive high schools in Philadelphia—Germantown, Gratz, Abraham Lincoln, and William Penn. Like many of the other PAHSCI schools, these four are identified as “high risk.” Students in these schools perform poorly on the statewide subject assessments (PSSA), the schools do not meet AYP targets, and they all have high dropout rates.

PAHSCI schools literacy and mathematics coaches are fully funded by the Initiative. The coaches provide an array of services to individual classroom teachers, and to groups of teachers in the form of professional development. The coaches are themselves supported

¹ For a complete description of PAHSCI and the coaching framework, see www.pacoaching.org
by mentors who bring additional subject matter expertise and support to both school leaders and classroom teachers. In addition, coaches and teachers at PAHSCI schools participate in training provided by the Pennsylvania Literacy Network, assuring a strong integrative literacy foundation to the Initiative.²

There are many ways in which PAHSCI influences schools and teachers, and these are now well documented.³ What remains is an overwhelming interest in the contribution of one-on-one coaching to improving student outcomes. Using data made available for secondary analysis, this paper examines evidence of student progress on the Philadelphia Benchmark Assessment in the classrooms of coached teachers in the four PAHSCI schools.

The data for this study were made available by the School District of Philadelphia (SDP). SDP had agreed to cooperate with PAHSCI on research subjects of mutual interest. Benchmark assessment data for 9th and 10th grade students were provided by SDP on four subjects: Algebra I and Geometry, and English 1 and English 2. Each of the four participating schools (“treatment” schools) was matched with a Philadelphia high school that was similar in its demographics and student performance on the PSSA, but did not offer PAHSCI coaching (“control” schools). Characteristics comprising the match of each treatment and control school included total enrollment, enrollment by ethnicity, percentage of students receiving free or reduced cost lunch, and percentage of students scoring at basic level or below on the statewide 11th grade reading and math assessments (PSSA).

² PLN is a comprehensive professional development/curricular enhancement program based in the Graduate School of Education at the University of Pennsylvania. PLN provides on-site credit-bearing or non-credit bearing seminars, workshops and mentoring programs in the areas of reading, writing, mathematics, science, and technology since 1981.
The Schools

The four treatment schools have enrollments ranging from about 1000 to nearly 2000 students. Three of the four—Abraham Lincoln is the exception—are over 90 percent African American. At all four schools over 60 percent of students receive free or reduced price lunch. At two of the schools nearly 80 percent of students scored “below basic” on the PSSA 11th grade reading assessment, at one school the figure was nearly 60 percent, and at the other school, nearly 70 percent. At two of the schools over 60 percent of students scored “below basic” on the PSSA 11th grade math assessment, and at the other two, over 60 percent scored “below basic.” Each PAHSCI school was matched with a control school on these variables. On most variables the matches were quite close.

The Benchmark Assessments

The Philadelphia Benchmark tests are low stakes interim tests aligned to the Pennsylvania academic standards. They are intended to assess student skill acquisition on discrete components of the curriculum in a number of subjects at ten week intervals through the year. The tests are given at many grade levels including 9, 10, and 11.⁴ Results are intended to help schools and teachers calibrate instruction to better prepare students for the statewide PSSA exams.

Organization of the Data

Working with the School District of Philadelphia (SDP) to carefully protect the identities of all concerned, SDP provided files comprising all students enrolled in the 9th and 10th grade in the four treatment and the four control schools. Student demographics, school attendance, 8th grade PSSA results, grades and more were provided in a second file. A third file of teachers linked each student to the teacher they took each class from; and a fourth file provided benchmark scores for each of the three test administrations in English and math during the 2006-07 school year. A final file identified each treatment school

teacher’s level of participation in one-on-one PAHSCI coaching. Teachers were scaled as high, meaning frequently coached one-on-one, low meaning occasionally coached, and uncoached. An elaborate process was used to link the files and create a data set enabling a school-by-school, grade by grade, subject by subject analysis of student performance on the Benchmark assessments for the year.

*Students included in the Analysis*

We looked at April, 2007 enrollment in the treatment and control high schools for 9th and 10th grades. Overall, there were 3,565 9th graders and 2,817 10th graders in the treatment schools. Our analysis sample, which required enough information to identify student characteristics, teacher participation in the coaching initiative, and the correct classes, contained 2,031 9th graders and 1,620 10th graders or about 58 percent of the total population across the four treatment schools. In order to focus on students most likely to have benefited from the efforts of their coached teachers, we limited the analysis to students who took all three benchmark tests, this as a proxy for regular school attendance. In the treatment and control high schools under 75 percent of students actually presented themselves for each benchmark exam. The analytical base was further reduced in specific subject analyses, since even fewer students in a benchmark-eligible class took all three benchmark tests.

*The Analysis Model*

We used several different statistical methods to examine the association between having a teacher who participated in the coaching initiative or having a teacher who did not, and student scores on each of the three benchmark tests in each subject. We controlled for students’ 8th grade PSSA scores, a proxy for academic performance. As noted above, we also limited the analysis to students who took the benchmark tests each of the three times it was administered, this as a way of capturing those students who were most likely to have attended school on a regular basis.
We calculated simple descriptive statistics and examined differences in means (t-tests and analysis of variance). The primary focus was on estimating linear regression models between test scores and teacher participation, to which additional control variables were added. See Appendix A for a general description of the model.

School by School and Subject Matter Comparisons—Results for Schools

These data were gathered retrospectively, that is, after the 2006-07 school year. As such, we tried to control for as many differences among students as we could. We paired treatment and control schools (i.e. those with and without coaching) according to demographics and similarity of 8th grade PSSA scores in reading and mathematics as noted above. Separate regression models were tested for English and mathematics classes at 9th and 10th grade, for each of the three benchmark exams. There was variation by school and by subject.

William Penn

English 1 and Algebra 1 (9th grade courses)

Students of coached teachers at the treatment school systematically outperformed students of uncoached teachers at the comparison school in both subjects.

   In English 1 students of teachers regularly coached one-on-one, outscored students of uncoached teachers at the comparison school on two of the three benchmark administrations--on average 7 points higher on the second administration, and 4 points higher on the third administration.

   In Algebra 1, students at William Penn scored on average 3 points higher on the first administration, 2 points higher on the second, and 9 points higher on the third administration, although only the last difference met conventional levels of statistical significance.
English 2 and Geometry (10th grade courses)

In English 2, students of teachers regularly coached one-on-one, scored on average 2 points higher then the students of uncoached comparison school teachers on the first administration, nearly 2 points higher on the second, and 6 points higher on the third, although none of these differences met conventional levels of statistical significance.

In Geometry, students of teachers regularly coached one-on-one, scored on average 3 points higher then the students of uncoached comparison school teachers on the first administration, 3 points higher on the second, and 12 points higher on the third.

Gratz High School

English 1 and Algebra (9th Grade courses)

In English 1, students of teachers regularly coached one-on-one, scored on average 1 point higher then the students of uncoached comparison school teachers at the first administration, and 3 points higher on the second administration and about the same as comparison school students on the third administration, although none of these differences met conventional levels of statistical significance.

In Algebra 1, there were no positive differences between Gratz students of teachers who were one-on-one coached and students at the comparison school.

English 2 and Geometry (10th grade courses)

In English 2, students of teachers regularly coached one-on-one, scored on average 7 points higher then the students of uncoached comparison school teachers on the first administration, 5 points higher on the second, and 5 points higher on the third.
In Geometry, students of teachers regularly coached one-on-one, scored on average 8 points higher than the students of uncoached comparison school teachers on the first administration, 4 points higher on the second, and 7 points higher on the third.

These two high schools showed the most promising results. Results for the two other coached schools did not show positive patterns either by grade or by subject measured against their comparison schools. At these schools rather few students took all three of the Benchmark tests. The small population made the analytics extremely problematic.

A Note on Student Engagement

The theory of change that underlies PAHSCI basically argues that coaches help teachers improve their instructional practice; that the improved instructional practice engages students in productive ways; and that engaged students are more involved in the life of the class and perform better on assessments of learning and achievement. Student engagement is often difficult to measure. One proxy is attendance. In that respect, it is interesting to note that students of coached teachers at William Penn, attended school, on average 7.5 days more than their comparison school counterparts; and students at Gratz attended school on average 17.3 days more than their comparison counterparts. These differences are material. Given a school year of approximately 180 days, the Gratz students were at school almost 10 percent more than their comparison school peers, and William Penn students were at school over 4 percent more than their comparison school peers. In this sense, the data suggest that this link to student engagement may be an important PAHSCI outcome in its own right, and an indicator of improvements in teacher quality associated with one-on-one coaching.
Understanding the Findings

There are many considerations to consider.

First, it is possible that the coaches at William Penn and Gratz were better able to gain the support of their teachers, and create a professional development climate that helped to change instructional practice. How much that is a function of the coaches’ skill, teacher receptivity to coaching, or some combination of both, along with support from school leaders, is not possible to say. The requisite combination of coaching skills and contextual enabling conditions may have been present in the two schools that showed evidence of change but not in the other treatment schools.

Second, it may be that there was varying coach capacity across schools, grade levels, or across subject matters. Perhaps these two treatment schools, by chance, had coaches who were better able to work effectively with teachers.

Third, even with careful attention to creating a strong match between coached schools and the comparison schools, there was still substantial variation in student characteristics and no control for contextual differences among the schools. This variation may have had analytical implications. Similarly, other programs and reforms may have overwhelmed any the specific gains related to coaching at treatment schools.

Finally, the findings may not have been statistically significant because so few cases were available in each regression when controlling for multiple factors including student 8th grade PSSA score, student attendance, student participation in all three of the Benchmark tests, subject matter, and teacher’s participation in one-on-one coaching. In a few cases, for a few subjects, the N’s fell below 100.
Where too from Here?

Taken together, these results are suggestive of student learning gains that may be associated with PAHSCI. A more systematic, carefully controlled outcomes-related study is necessary to truly understand how, why, and under what conditions coaching yields positive results for students.
Appendix A
The Model

The baseline model for this analysis was:

\[ Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \]

Where:

- \( Y_i \) = score on benchmark test “i”, where \( i = 1, 2, \) or \( 3 \)
- \( X_1 \) = dummy variable indicating student had a treatment school teacher who was not a participant in the coaching initiative.
- \( X_2 \) = dummy variable indicating student had a treatment school teacher with a low level of participation in the initiative.
- \( X_3 \) = dummy variable indicating student had a treatment school teacher with a high level of participation.

The reference category was having a subject matter teacher in one of the control schools.

- \( \beta_0 \) = constant term
- \( \beta_1, \beta_2, \beta_3 \) = are unknown coefficients to be estimated. Each coefficient reflects the average difference in benchmark scores between control group students and the particular group of students represented by the dummy variables \( X_1-X_3 \)
- \( \epsilon \) = a normally distributed error term with mean zero and constant variance

Our analyses progressed from this relatively simple model to include controls for pre-existing differences in student ability, as measured by 8th grade PSSA scores in math and reading (depending on the which benchmark subject was being analyzed). Other analyses added student attendance (measured in days) to represent dosage or exposure. In addition, we also ran analyses separately by each of the four pairs of treatment and control schools, i.e., the models were limited to include only students who attended the specific treatment/control group, say Germantown and West Philadelphia, and then for William Penn and Bartram, and so on). Finally, we also ran some models limiting the sample to students who had taken all three benchmark tests.