

Instructional Coaching and Student Outcomes

Findings from a Three Year Pilot Study

Elliott A. Medrich, Principal Investigator

Robert Fitzgerald, Statistician

Paul Skomsvold, Research Assistant

Prepared under contract to

The Pennsylvania Institute for Instructional Coaching

MPR Associates, Inc.

2150 Shattuck Avenue, Suite 800

Berkeley, CA 94704

Contact

Elliott A. Medrich

emedrich@mprinc.com

510-849-4942

January 2013



innovative research • practical solutions

Abstract

Teachers and teaching are at the forefront of the school reform agenda. There is a consensus that in addition to better teacher preparation, a stronger curriculum, better diagnostic tools and assessments, and significant changes to accountability systems, effective professional development is one key to improving the quality of instruction in schools. Instructional coaching is one approach to professional development that is of particular interest in many schools and school districts across the country.

This study focuses on a coaching model designed by the Pennsylvania Institute for Instructional Coaching (PIIC), a project supported by the Annenberg Foundation and the Pennsylvania Department of Education (PDE). The PIIC model emphasizes the simultaneous use of four strategies: one-on-one teacher engagement; evidence-based literacy practices applied across the curriculum; data analytics; and reflection on practice.

Taking advantage of an unusual opportunity, this study broadly glimpsed the effects of coaching, looking for relationships among coaching, teaching, and student outcomes as reflected in student performance on statewide assessments in reading (PSSA). We were able to identify a treatment school where all teachers, at all grade levels in the school were coached, and we were able to obtain individual level student standardized test results for students at that school. For comparative purposes, we also identified two schools, each with similar demographic and performance profiles to the treatment school, but without a coaching program.

We posed three research questions—one related exclusively to student outcomes on the reading PSSA within the treatment school where teachers were coached; and two involving multi-year comparisons of student outcomes between the treatment school and two comparison schools, where teachers were not coached. These were the research questions:

How have students at the treatment school performed on PSSA reading, as compared with predicted estimates from the state value added assessment (PVAAS), during the time the school had a full-time coach working with teachers at all grade levels?

From grade level to grade level, over a three year research cycle (pre-coaching, first year of coaching—the base year—and the following year), how do changes in the performance of students on PSSA reading in the treatment school compare with the performance of students in the comparison schools? Do incoming 3rd, 4th, and 5th graders, etc. show increasing proficiency in the treatment school where teachers were coached compared with students in the same grades in the comparison schools which did not have coached teachers?

For each grade cohort, over a three-year period, how do changes in the performance of students on PSSA reading at the treatment school vary as students progress from grade to grade, compared with the performance of students in the comparison schools?

Data were gathered over a period of three years from two sources—the Pennsylvania Value Added Assessment System (PVAAS) and the Pennsylvania Standardized Student Assessment (PSSA). Individual unit record data were available for students in the treatment school and aggregate data for students in the comparison schools.

On all three research questions, there was evidence to improvement in PSSA reading results at the treatment school at rates often exceeding results at the comparison schools. This was true for some grade levels more than for others and for some student cohorts, including economically disadvantaged students, more than for others. But taken as a whole, the results offer some positive indications as to the efficacy of instructional coaching as a professional development initiative.

Instructional coaching is not a practice that can be viewed in isolation. At its best, it supports a quality instructional program in a school. The strategic delivery of coaching over time can have an impact on teachers' instructional practice and, in turn, this can help change the nature of teaching in ways that lead to improved student learning.

Contents

Abstract	ii
List of Figures	v
Introduction	6
A Sense of the Literature	7
Coaching and Student Outcomes	8
The Study Setting: Treatment and Comparison Schools.....	9
Research Questions and Analysis.....	11
Limitations of the Research Design.....	11
The First Research Question: Outcomes at the Treatment School	12
The Second Research Question: Grade-Level Results for Treatment and Comparison Schools.....	14
The Third Research Question: Cohort Results for Treatment and Comparison Schools.....	18
Discussion	21
Final Consideration.....	22
Appendix A—Intensity of Coaching at the Treatment School.....	23
Appendix B—Technical Notes	25
Statistical Tests	25
Data Sources.....	25
Number of Students Scoring Proficient or Advanced on the PSSA, by Year, Grade, and Cohort—Treatment and Comparison Schools.....	26
Appendix C—Selected Studies of Instructional Coaching Outcomes.....	29
Appendix D—Characteristics of Treatment and Comparison Schools.....	31

List of Figures

Figure 1:	Percentage of students with a one-year PVAAS projection of 50 percent or greater probability of scoring proficient or higher on PSSA reading, and the percentage of those students who subsequently scored proficient or advanced on the PSSA the following year, by grade level.....	13
Figure 2:	Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3, 2009–12	14
Figure 3:	Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4, 2009–12	15
Figure 4:	Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5, 2009–12	15
Figure 5:	Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3, 2009–12.....	16
Figure 6:	Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4, 2009–12.....	17
Figure 7:	Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5, 2009–12.....	17
Figure 8:	Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5 th through 8 th grade cohort	19
Figure 9:	Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 6 th through 8 th grade cohort	19
Figure 10:	Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 4 th through 7 th grade cohort	20
Figure 11:	Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5 th through 8 th grade cohort	20

Introduction

Teachers and teaching are at the forefront of the school reform agenda. There is a consensus that in addition to better teacher preparation, a stronger curriculum, better diagnostic tools and assessments, and significant changes to accountability systems, effective professional development is one key to improving the quality of instruction in schools. That said, evidence suggests that traditional professional development, often characterized by “once-and-done” occasional lectures and workshops without follow up, is mostly of little value or at best only marginally useful to teachers. Extensive and extended job-embedded opportunities to build teaching skills with non-evaluative support, over time, are the currency of the realm today.

Broadly speaking, most professional development is intended to assist teachers in helping their students learn. If this were not the objective, why else would we invest scarce resources—time and money—in the myriad activities that make up the professional development enterprise? Indeed, professional development is but one component of any strategy to improve classroom teaching and student learning. Successful professional development should be tied to schoolwide learning strategies and an excellent curriculum.

Instructional coaching is one approach to professional development that is of particular interest in many schools and school districts across the country. It follows a long, time-honored and well-respected tradition—from executive coaching in the business sector, to team and personal coaching in athletics, to voice, dialect, and instrumental music coaching in the arts. In fact, coaching is well-recognized for its contribution to nearly every sphere of professional life. In schools, the concept of instructional coaching is equally appealing and for many of the same reasons: it can represent a means of personalizing the way teachers as professionals gain new knowledge and improve their skills. As in other sectors, coaching in schools can take many forms. There is no single definition of what instructional coaching is and what it is not. Hence, any inquiry into the impact of instructional coaching on teachers, and on student learning, must begin with a clear description of the intervention, its objectives, and its intended outcomes.

This study focuses on a coaching model designed by the Pennsylvania Institute for Instructional Coaching (PIIC),¹ a project supported by the Annenberg Foundation and the Pennsylvania Department of Education (PDE). The PIIC model emphasizes the simultaneous use of four strategies: one-on-one teacher engagement; evidence-based literacy practices applied across the curriculum; data analytics; and reflection on practice.

¹ Pennsylvania Institute for Instructional Coaching, www.pacoaching.org.

PIIC instructional coaches are trained to engage teachers one-on-one, side-by-side. They employ a “BDA” strategy of consultation (“before, during, after”) when working with teachers.² Coaches are supported by mentors (a “coach’s coach”), who provide training to ensure that the coaches—not just the coached teachers—are continually building their own skills (a unique feature of the PIIC model). Further, PIIC coaches promote the use of evidence-based literacy practices and research-based instructional techniques utilizing a well-regarded literacy framework as a central professional development component. Coaches enjoy strong professional support through ongoing statewide gatherings designed to sharpen their skill sets.

Not all coaching is alike. And not all instructional coaching emphasizes the set of four strategies mentioned above.

A Sense of the Literature

There are many studies of instructional coaching, but by and large, they focus on implementation and description of coaching strategies. Rather few examine outcomes for teachers or students (a sampling of this research is referenced in Appendix C).

While many feel that instructional coaching represents an attractive approach to delivering professional development, with a presumed impact on teachers’ instructional practice and student learning, there is relatively little research that shows the *effects* of coaching on student outcomes.³ This is not entirely surprising. Linking coaching to instructional practice *and* to student outcomes is a challenge—and finding effects of sufficient size to attribute a direct association among them is difficult at best. As noted in a recent, extensive review of two instructional coaching models, Janet Quint writes generally about coaching as professional development:

The theory (of action) hypothesizes that professional development will improve both teachers’ content knowledge and their instructional practices. As a result of improved instruction, students’ achievement will also improve, as measured by scores on tests measuring their...skills.⁴

² “BDA: stands for “Before-During-After.” It is a commonly used coaching strategy. For details, see Eisenberg, E. (2011, July 25). *Coaching 101*. Presentation at a professional development conference sponsored by the Pennsylvania Institute for Instructional Coaching.

³ For example see Knight, J., & Cornet, J. (n.d.). *Studying the impact of instructional coaching*. Lawrence, KS: Center for Research on Learning, University of Kansas.

⁴ Quint, J. (2011, July). *Professional development for teachers: What two rigorous studies tell us*. New York: MDRC, p. 3.

Quint concludes that while the theory of action is reasonable enough, changes in teacher knowledge and practice would have to be very substantial to “move the needle on student achievement significantly.”⁵

Quint’s study highlights the daunting challenge associated with research that attempts to associate each component part of the equation—coaching *and* instructional practice *and* student learning—and it is not surprising that this relationship is not well-researched. But ultimately instructional coaching must show evidence that there are benefits and positive outcomes for students. That is what policymakers and practitioners need and want to see.

Coaching and Student Outcomes

Arguably, every intervention in a school should directly or indirectly help improve outcomes for students. Coaching is no exception, but the relationship is complex, difficult to measure, and not necessarily associated with simple metrics.

Schools are complex organizations, and coaching at its best is just one element in a larger instructional framework. Coaches work with many teachers, and in most settings, even the most intensive coaching experience for teachers is measured in modest doses: perhaps a few hours per week or per month. To the extent that this is the case, what outcomes are reasonable to expect from a coaching intervention? This is an important question, because attempts to measure the impact of coaching on teachers and students must be realistically aligned with the actualities of the intervention.

Some coaching is focused on the micro-level. As an “ideal” example, a teacher is coached on improving a particular class lesson plan. This is “ideal” in that there are no exogenous factors affecting student scores—the entire teaching and learning experience is controlled from start to finish. It is offered to all students in the same way. The teacher learns from the coach and uses the new skills in her or his classroom. The students show gains measured pre- and post-lesson, and these gains can be linked back to the teacher’s time with the coach on that specific lesson—an excellent outcome. Hopefully, what the teacher has learned will also be applied in other ways, in other lessons. So, one could study the effects of coaching based on particular coach–teacher interactions and related student outcomes (pre- and post-coaching assessments), and this would be a valuable way of viewing the incremental effects of coaching on teachers and students.

⁵ Quint, p. 22.

It is important to bear in mind that instructional coaching is, in the first instance, a process designed to help teachers improve their practice. While we expect that improvements in instructional practice will lead to improvements in student learning, from a research perspective, the most significant impact of instructional coaching is likely to be reflected in changes in how teachers teach—teachers are, in fact, the target of the intervention.

That said, because there is an expectation that reforms and interventions in schools will ultimately improve student learning *writ large*, findings of student success at the micro-level are undervalued, and most policymakers want to see results that can be associated with student performance on statewide annual standardized examinations—i.e., the metric “that matters.” Coaching may affect student outcomes, but not in a way that is evident in student annual standardized test results.

In the best of circumstances, the answer to the questions “how have students benefited from coaches supporting teachers, and what would have happened if their teachers had not experienced the intervention” would be addressed through a randomized controlled trial (RCT). RCTs, or experiments, are generally considered to be the gold standard in the analysis of educational interventions. Under most conditions, random assignment ensures that the group receiving the intervention is the same as the group not receiving the intervention.⁶ RCTs are often infeasible, however, because of resource limitations and ethical concerns. In this case, there was no possibility of conducting a random assignment study. Neither the modest research budget nor the structure of the coaching program itself made that possible.

The Study Setting: Treatment and Comparison Schools

Taking advantage of an unusual opportunity, this study has been able to take a broad glimpse at the effects of coaching, looking for relationships among coaching, teaching, and student outcomes as reflected in student performance on statewide assessments. We were able to identify a school where all teachers, at all grade levels in the school, were coached, and we were able to obtain data about individual student outcomes at that school. It is rarely the case that every teacher in a school is coached the same way on a regular basis

⁶ Random assignment is expected to produce treatment and control groups that are similar on both observable and unobservable characteristics (the latter might include “motivation to learn” or other personality characteristics associated with student performance). In an ideal experiment, teachers would be randomly assigned to receive coaching and students would be randomly assigned to classrooms with teachers who are or are not receiving coaching.

(see Appendix A). The treatment school used coaching as a whole school strategic reform—a commitment to instructional coaching permeated the school culture—and thus we were able to explore the effects of coaching in a “full immersion” environment.

The treatment school itself is in a relatively low income area of the Commonwealth—62 percent of students were free lunch eligible. It was ethnically homogeneous. There were very few minority students. Among all students in grades 3–6, 58 percent tested below basic or basic on the state’s (Pennsylvania System of School Assessment) PSSA exam in 2009–10,⁷ the year that coaching was initiated. There was one teacher per grade level and one coach who worked with every teacher at every grade level. The coach had been at the school for eight years and enjoyed widespread support from administrators and teachers alike. The coach was trained by PIIC and had participated in other coaching initiatives in the past. The school’s coach received ongoing support from the PIIC area mentor.

At the treatment school there was *fidelity* to the PIIC model, as measured by the PIIC organizing principles. Because the school had just one coach, there was no issue concerning uniformity of practice—all teachers were coached in the same way. That said, not all teachers in the school were coached to the same level of intensity. Using a rubric to define the *intensity* of coaching given to each teacher in each grade, we classified the level of coaching by individual teacher. Given that data were gathered over a three-year period and staff turnover was very low, almost all students were in the class of at least one teacher who received a high level of coaching during the data collection. (For a description of the coach–teacher coaching classification rubric, see Appendix A.)

The treatment school principal and the regional administration were committed to full participation in the instructional coaching initiative. A considerable effort was made to maintain *fidelity* to the PIIC model, with a strong emphasis on supporting the reading and literacy curriculum. Further, the school principal had served as an instructional coach in the past, was familiar with the related processes, and provided extensive within-school support to the instructional coach. Coaching was adopted with conviction and received with enthusiasm by most of the teachers.

For comparison, with the help of the SAS Institute—which designed and manages the Pennsylvania Value Added Assessment System (PVAAS)—two Pennsylvania schools were carefully matched with the treatment school. The comparison schools were similar to the treatment school along four dimensions: grade configuration; enrollment; percentage of students who were economically disadvantaged; and most important, similarity in average

⁷ For a description of the PSSA, its component parts, and scaling, see Pennsylvania Department of Education, *2011–12 PSSA/PSSA-M Assessment Handbook*, available at http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_system_of_school_assessment/%28pssa%29/8757/resource_materials/507610.

student growth on the state's value-added performance measures during the year prior to the coaching intervention (see Appendix D). In making the match, no information about teachers' experience or expertise was available for the comparison schools.

The K–8 treatment school was located on the outskirts of a metropolitan area. Comparison school A was located in a rural area, and Comparison school B was in an urban district. Neither comparison school supported any form of instructional coaching.

Research Questions and Analysis

This pilot study explores research questions, as described in detail below, examining several kinds of student outcomes—one related exclusively to student outcomes within the treatment school, and two involving multi-year comparisons of student outcomes between the treatment school and the comparison schools.

- *Within the treatment school:* an analysis of projected student PSSA scores as compared with actual student PSSA scores (using data from the Pennsylvania Value Added Assessment System—PVAAS) by grade, gathered over a three-year period;
- *Between treatment and comparison schools:* comparisons of student performance on the PSSA by grade and by student cohort over time.

Data from the treatment school included both unit records and aggregate-level information on students by grade. Data for the comparison schools were only available at the aggregate level by grade.

Limitations of the Research Design

There were significant constraints on data availability and budget that affected the analytic approach and the overall research design. Among these were the following:

- Unit record data for students in the comparison schools were not available, so we could not examine how comparable students in the treatment and comparison schools progressed nor could we control for pre-existing differences in student achievement. Although we know that the grade-specific cohorts were relatively stable from year to year in the treatment school, we do not know that the same was true for the comparison school cohorts.

- While we would have preferred to conduct a more rigorous non-experimental study using propensity score matching of treatment and control school students, for example, this was not possible.
- We did not have information on the characteristics of teachers in the comparison schools.
- The number of students in each school overall and in each grade within schools was small, so the power to detect statistically significant effects of coaching is also small. In addition, small changes in the number of students rated proficient or advanced on the PSSA may have large effects on proficiency rates. As appropriate, measures of statistical significance were calculated for the year-to-year changes to assist readers in assessing these changes.
- Finally, the analysis does not use a multi-level approach and there are no adjustments to standard errors for clustering.

The First Research Question: Outcomes at the Treatment School

How have students at the treatment school performed on PSSA reading, as compared with PVAAS estimates, during the time the school had a full-time coach working with teachers at all grade levels? To address this question, we used unit record PVAAS data prepared by the SAS Institute, estimating the probability that a given student would achieve a proficient or advanced score on the PSSA reading examination the *following year*. PVAAS is a well-known and widely respected estimation model developed and designed by the SAS Institute for the Commonwealth of Pennsylvania.

PVAAS is a statistical analysis of PSSA data and provides districts and their schools with progress data to add to achievement data. This new lens of measuring student learning provides educators with valuable information to ensure they are meeting the academic needs of cohorts of students, as well as individual students.⁸

So the first question asks whether the students met or exceeded expectations, as defined by a comparison of students' expected performance on the statewide PSSA reading exam as estimated by PVAAS and their actual performance. It is intended to show progress at the

⁸ For a more complete description of PVAAS and the PA system, see http://www.portal.state.pa.us/portal/server.pt/community/pa_value-added_assessment_system_%28pvaas%29/8751/introduction_to_pvaas/507600.

treatment school against the PVAAS measure. This question does not refer to the comparison schools, which may or may not have had similar outcomes against PVAAS expectations.

From PVAAS data, we were able to ascertain the proportions of students at the treatment school who were expected to meet or exceed an estimated PSSA result for the following year (i.e., PVAAS takes the actual growth estimate for one year and projects the students expected outcome—below basic, basic, proficient, or advanced—for the following year).

We analyzed students' actual performance against expectations. An example for the school years 2010–11 and 2011–12 is offered below (figure 1).

Figure 1. Percentage of students with a one-year PVAAS projection of 50 percent or greater probability of scoring proficient or higher on PSSA reading, and the percentage of those students who subsequently scored proficient or advanced on the PSSA the following year, by grade level

2009–10 Cohort	2010–11 Percent Proficient or Advanced	2011–12 Percent Proficient or Advanced
Grade 3	87.5	100.0
Grade 4	†	†
Grade 5	83.3	85.7
Grade 6	91.7	100.0
Grade 7	78.9	–
Grade 8	–	–

– Data missing or not applicable.

† Suppressed due to cell size.

SOURCE: Based on individual-level estimates by PVAAS and actual results on the PSSA for students by grade at the treatment school.

The treatment school results are quite positive, and even for students with a lower probability of scoring proficient or advanced on the PSSA reading, the school's results have improved from year to year. On this first measure, it seems that the coaching environment has helped the school bolster student performance on the PSSA and meet or exceed expectations, something they were not able to do before coaching was instituted as a primary mode of professional development.⁹

⁹ PVAAS data for the comparison schools were not made available, so it is not known whether these schools shared a similar outcome. Further, it is not known whether PVAAS systematically underestimates aggregate performance on the PSSA for the following year across schools throughout the Commonwealth.

The Second Research Question: Grade-Level Results for Treatment and Comparison Schools

From grade level to grade level, over a four year research cycle (pre-coaching, first year of coaching—the base year, and following year), how do changes in the performance of students on PSSA reading in the treatment school compare with the performance of students in the comparison schools? Do incoming 3rd, 4th, and 5th graders, etc. show increasing proficiency in the treatment school where teachers were coached compared with students in the same grades in the comparison schools which did not have coached teachers?

Although some volatility is evident, students in grades 3, 4, and 5 of the treatment school showed a year-to-year positive trend in their PSSA scores (base year—the year prior to introduction of the coaching initiative—followed by three years of coaching). The comparison schools showed a downward trend in these three grades over the four-year research cycle.

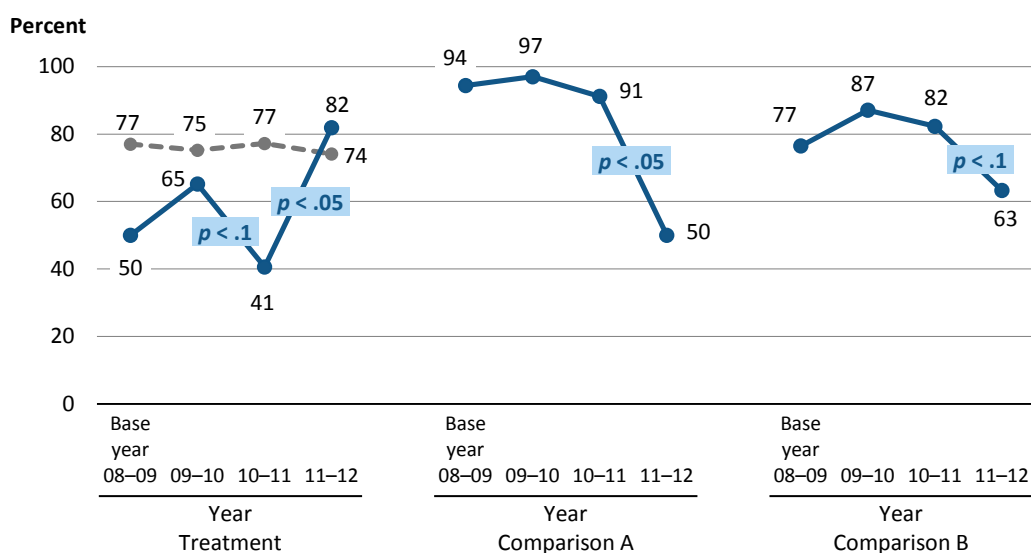
Results for grades 3, 4, and 5 are illustrated below for the treatment and comparison schools (figures 2–4).

Key to Reading the Figures

2008–09: Base year (pre-coaching)
2009–10: First year of coaching initiative
2010–11: Second year of coaching initiative
2011–12: Third year of coaching initiative

- Dashed line represents year-to-year statewide reading PSSA results for proficient and advanced performance.
- Solid line represents year-to-year *treatment* and *comparison* school reading PSSA results for proficient and advanced performance.
- Results of significance tests shown on the figure for year-to-year change where $p < .1$ and $p < .05$.

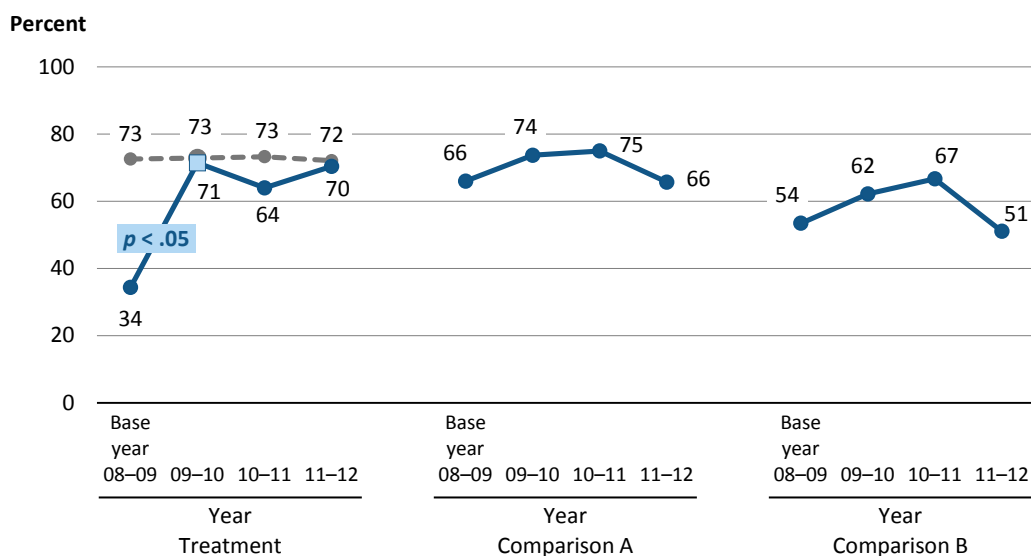
Figure 2: Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3, 2009–12



NOTE: The dashed line represents the statewide percentage of all students rated proficient and advanced in grade 3 2009–12.

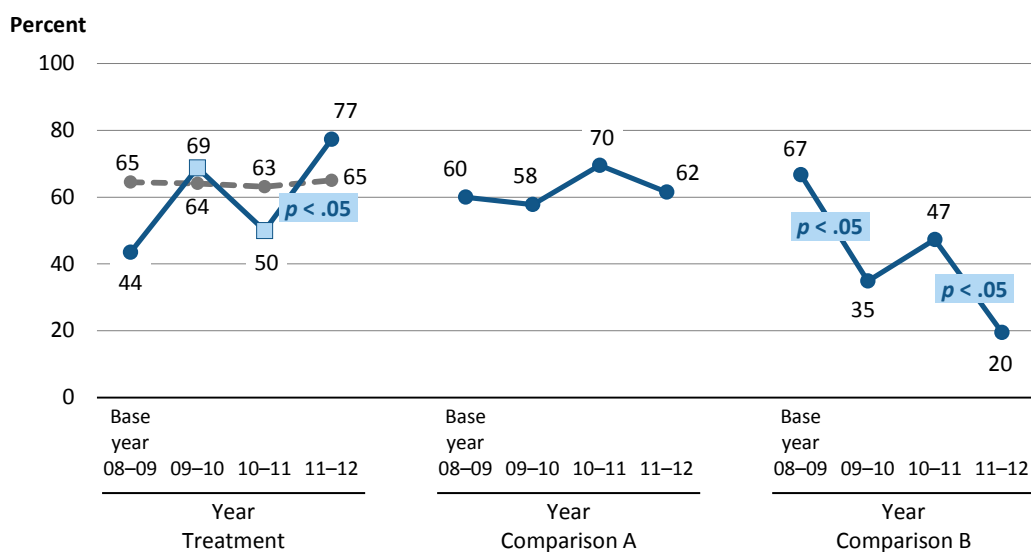
SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

Figure 3: Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4, 2009–12



NOTE: The dashed line represents the statewide percentage of all students rated proficient and advanced in grade 4 2009–12. For the treatment school, 2009–10, 4th-grade data (only) based on unit record results on the PSSA reading. SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

Figure 4: Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5, 2009–12

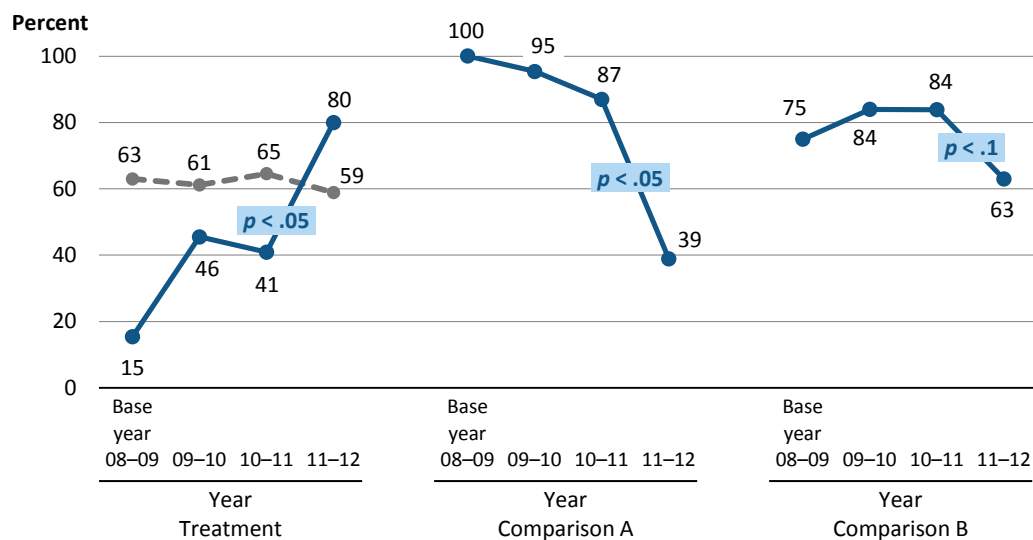


NOTE: The dashed line represents the statewide percentage of all students rated proficient and advanced in grade 5 2009–12. For the treatment school, 2009–10 and 2010–11 (only), 5th-grade data based on unit record results on the PSSA reading. SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

In grades 6 and 7, there were mixed results for students in both treatment and comparison schools, while students in grade 8 had positive results in both schools (not shown).

There are similar positive indications among 3rd-, 4th-, and 5th-grade economically disadvantaged students in the treatment school (figures 5–7).

Figure 5: Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3, 2009–12



NOTE: The dashed line represents the statewide percentage of economically disadvantaged students rated proficient and advanced in grade 3 2009–12.

SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

Figure 6: Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4, 2009–12

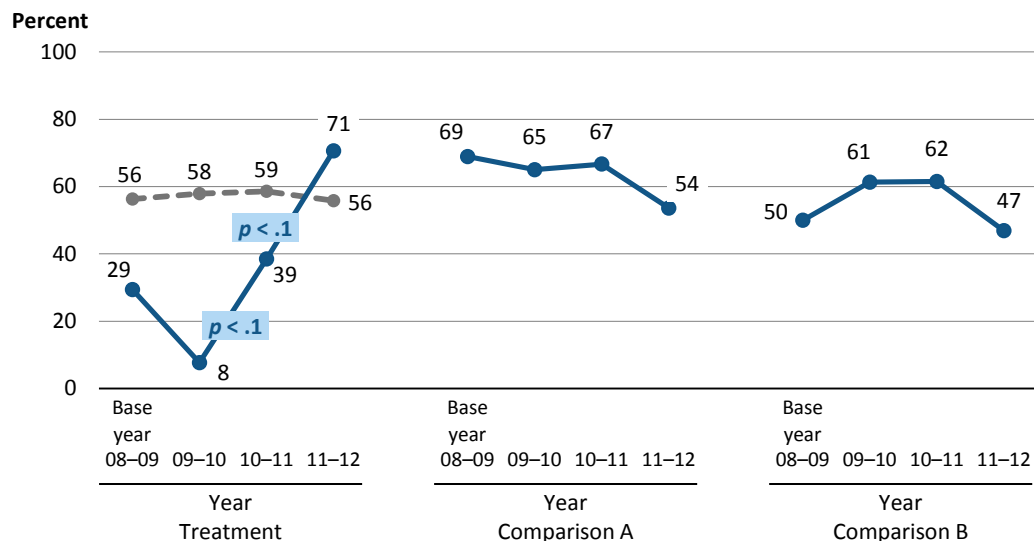
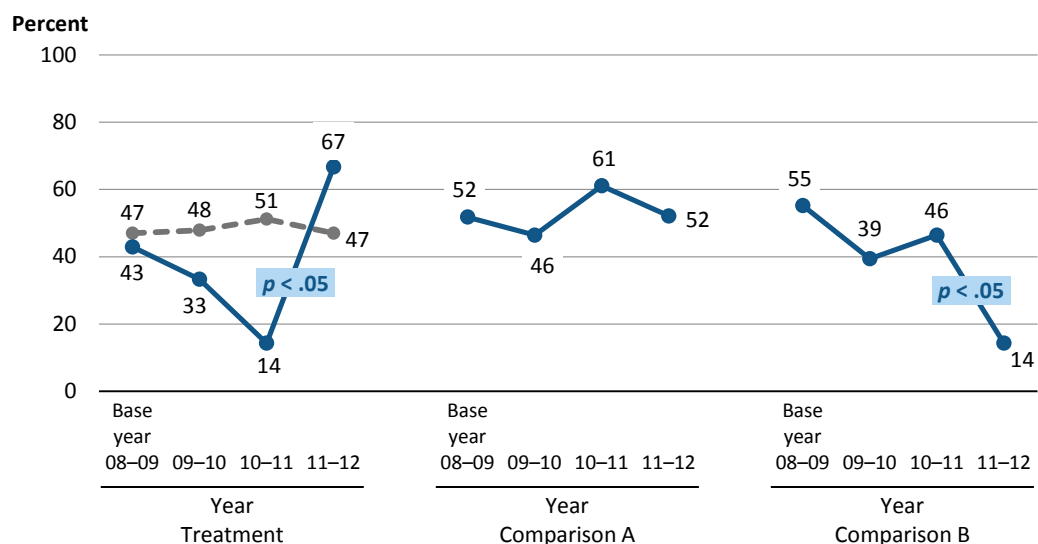


Figure 7: Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5, 2009–12



The positive indications for the lower grades are of interest, suggesting that the teachers in these classrooms became stronger over time, to the benefit of students. While one would hope to see a similar pattern at every grade level, as described in the discussion section below, there could be many explanations as to why this does not occur. Further, there was less teacher coaching at the upper grade levels, and that may have had some effect on the findings. That said, it is not entirely surprising to see an intervention that is more successful at some grade levels than at others.

The Third Research Question: Cohort Results for Treatment and Comparison Schools

For each grade cohort, over a three-year period, how do changes in the performance of students on PSSA reading at the treatment school vary as students progress from grade to grade compared with the performance of students in the comparison schools?

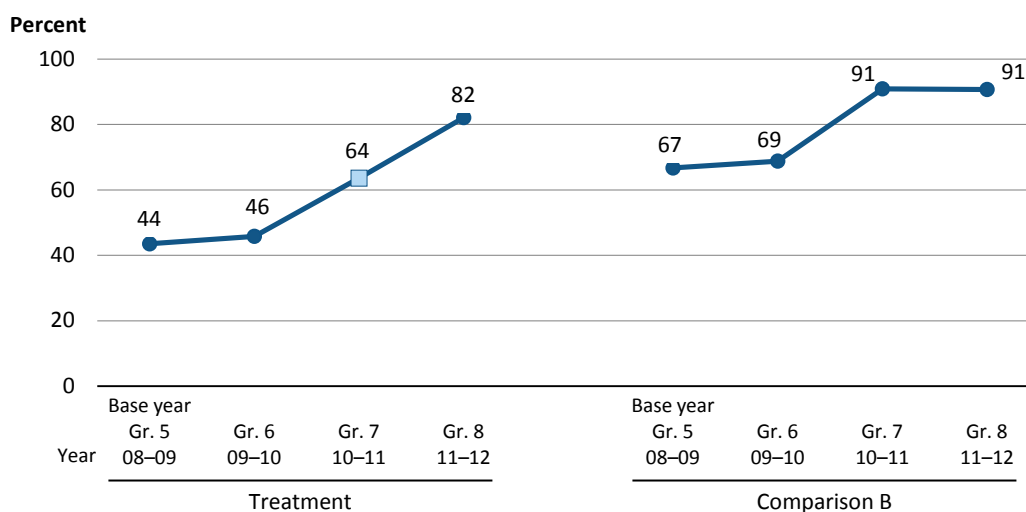
For this analysis, we looked at the PSSA reading results for students who entered a grade in the year before the coaching intervention began and followed their performance from grade to grade over the next three years, at both the treatment and the comparison schools. In other words, we followed the same students over a period of years.¹⁰

At the treatment school, the 5th-grade student cohort (over the course of 5th, 6th, 7th, and 8th grades) made particularly impressive gains in PSSA reading scores as compared with the same cohort at the comparison school (figures 8 and 9). And the cohort of 6th-grade students at the treatment school (over the course of 6th, 7th, and 8th grades) made significant gains compared with the same cohort at the comparison school.¹¹

¹⁰ Because we did not have unit record data for students in the control schools, we generally assumed that the results for starting 4th graders, for example, were reflected in the 5th grade the following year and that similarly 6th-grade results in the year after that were based on the same group of entering 4th graders two years previously. It is possible for the composition of the cohorts to vary as students entered and left the schools over time. This is a limitation of the data to which we had access. But as noted in Appendix B, the treatment school cohort remained stable with little attrition. Data are reported only for those students who remained at the treatment school for the entire three years of the intervention.

¹¹ Only one of the two comparison schools included grade 8.

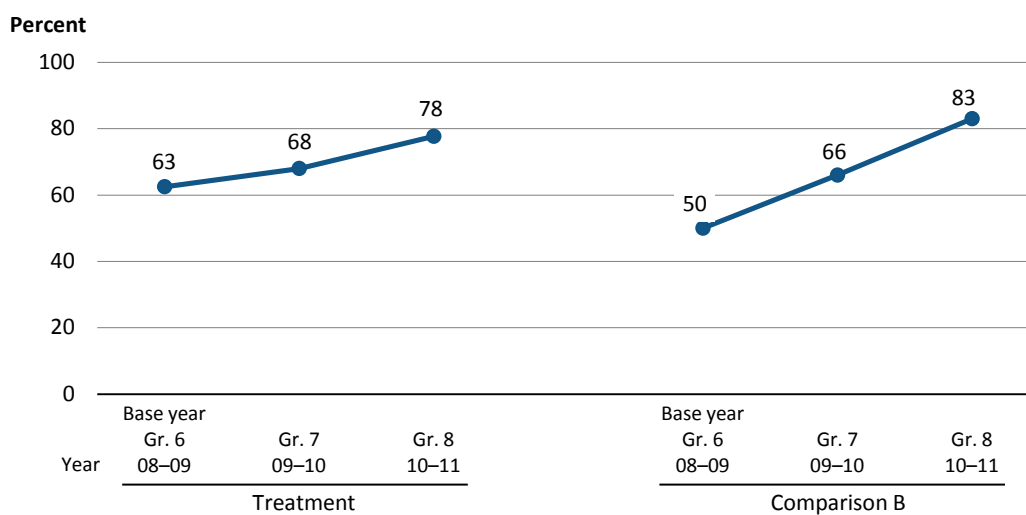
Figure 8: Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5th through 8th grade cohort



NOTE: Base year is year prior to introduction of coaching at treatment school.

SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

Figure 9: Percentage of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 6th through 8th grade cohort

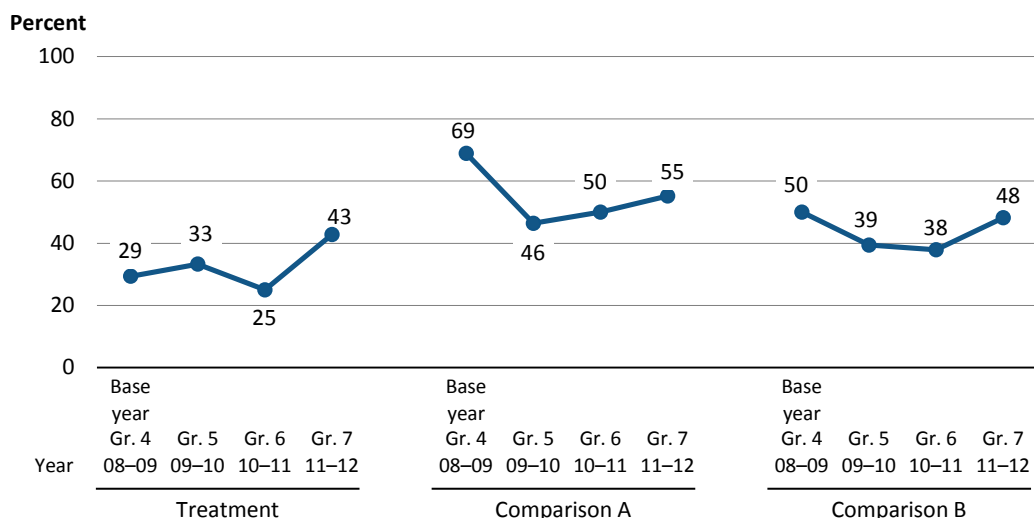


NOTE: Base year is year prior to introduction of coaching at treatment school. For the treatment school, 2010–11 (only), 7th-grade data based on unit record results on the PSSA reading.

SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011–12 and prior years.

In addition, cohorts of 4th- and 5th-grade students who were economically disadvantaged also showed favorable gains as compared with their counterparts at the comparison schools (figures 10 and 11).

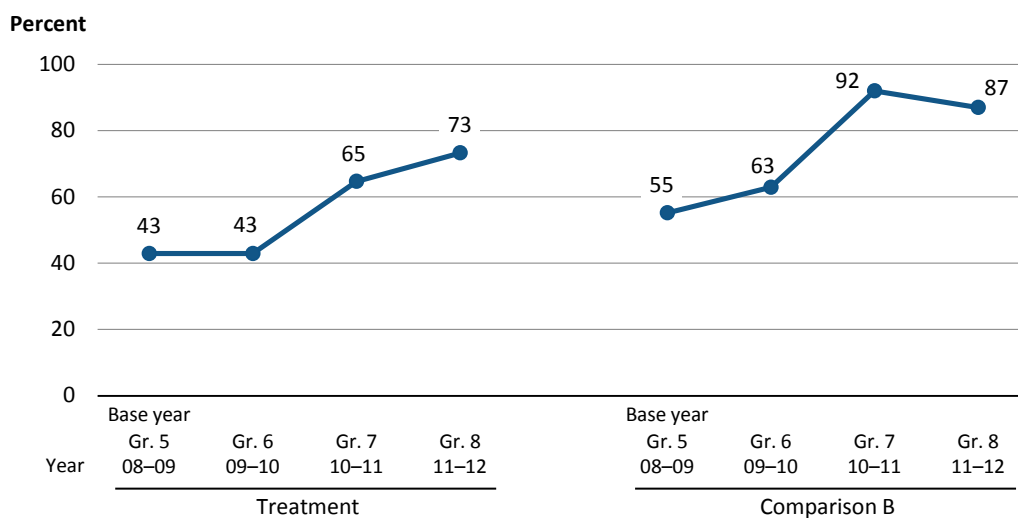
Figure 10: Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 4th through 7th grade cohort



NOTE: Base year is year prior to introduction of coaching at treatment school.

SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011-12 and prior years.

Figure 11: Percentage of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5th through 8th grade cohort



NOTE: Base year is year prior to introduction of coaching at treatment school.

SOURCE: http://www.education.state.pa.us/portal/server.pt/community/school_assessments/7442 for 2011-12 and prior years.

The results for economically disadvantaged students should be viewed with additional caution because their numbers were small, and data for fewer than 10 students were suppressed. However, the results also suggest some good progress at the treatment school, as compared with similar students in the comparison schools.

Discussion

Student outcomes are the product of many factors, of which instructional coaching at the treatment school is only one, so our conclusions about the effects of coaching should be viewed as tentative. That said, there are some positive indications that instructional coaching has provided an added dimension to the school's professional development program and that coaching affects teachers in ways that help them improve their instructional capacity.

While there were some positive indications that student PSSA reading results improved for some grade levels and for some student cohorts at the treatment school, this was not the case for all students at all grade levels or for all cohorts. Here are some considerations.

One might hypothesize that improvements in student learning at the treatment school may, in and of themselves, reflect superior teaching by certain teachers at certain grade levels. It may be true that some teachers are better than other teachers. That's a possibility, but at the same time, given that the same teachers have been at this school for years, if the teachers were very good, why didn't their students have better outcomes even before the coaching initiative began? That was not the case. But because coaching was adopted schoolwide, PSSA reading scores have improved over time at a faster rate than the state average, and at a faster rate than in the comparison schools. There is some evidence that instructional coaching has contributed to these gains.

This suggests a second consideration—momentum. It may well be the case that as teachers are coached from year to year, they become better at their craft (no matter how “good” they were before the coaching initiative began). As a result of their improved practice, their students achieve better than that same teacher's students achieved in prior years. That is to say, “all boats rise” when a school commits to coaching and sustains a high level commitment aimed at *all* teachers over a period of years. While the students of every teacher at every grade level did not advance to the same degree, the cohort data that shows gains for students beginning at the 5th grade and sustained over time is a decidedly good outcome.

This leads to a third consideration—momentum in student learning. Although these results are by no means definitive, nor are they equally positive for all cohorts, it may be that as

teachers improve their practice over time, and as students are exposed to teachers whose skill levels are improving from year to year, the student outcomes themselves improve for a particular student cohort over a period of years. Given where the treatment school started, student outcomes were extremely low, and given where the school is now, it seems reasonable to suspect that better teaching is leading to better outcomes in a cumulative fashion.

Again, it must be noted that this is a small study in an environment in which every teacher was coached and every student had the possibility of benefiting from whatever improvements in teaching might have occurred. Recognize that other things may have changed during this period, e.g., improved curriculum, better uses of data diagnostically, which may not be directly associated with coaching. *But even to the extent that this may be the case, the coach provided the school's primary professional support to teachers as they instituted these changes.* Without regard to results for each and every grade level and every student cohort, perhaps the most important point is that the coach provided the rigor, discipline, and professional “glue” that helped teachers improve their instructional capacity and implement changes in their practice. And student learning, as reflected in PSSA reading scores, improved and students exceeded PVAAS expectations. The level and direction of change at the treatment school should be the major takeaway from this pilot study.

Final Consideration

Instructional coaching is not a practice that can be viewed in isolation. At its best, it supports a quality instructional program in a school. It can, however, have a profound impact on teachers and, in turn, this can lead to improved student learning. The systematic delivery of coaching, over time, can help teachers improve their practice in innumerable ways and, thereby, change the nature of teaching and learning in the school environment.

Appendix A—Intensity of Coaching at the Treatment School

As noted in the text, understanding the degree to which coaches adhere to the coaching model (*fidelity*) is important, as is understanding the degree to which all teachers have similar kinds of support and professional development (*ubiquity*). A final consideration is the level (*intensity*) of coaching that a teacher receives. These three dimensions are critical, and they are closely related one to the other.

After exploring the fidelity and ubiquity of coaching at the treatment school and determining that the coaching model was closely followed and that all teachers were coached in the same way, teachers in the treatment school were scored on the intensity of coaching they received. While there are some general scales that address this issue,¹² there are considerations specific to the PIIC coaching model.

There were two particular considerations that figured into the teachers' intensity of coaching score for this research:

1. Participation in and knowledge of the literacy training program made available to PIIC coaches and their teachers.

PIIC coaches and some teachers receive formal training in a literacy strategy developed by the Penn Literacy Network (PLN) from the University of Pennsylvania.¹³ The classroom-based coursework involves participation in approximately 30 hours of classroom instruction and 30 hours of additional work with the coach.

2. Active engagement with the coach in the classroom on a regular basis (measured by frequency of visits); participating in the co-design of coaching sessions to meet the teachers' needs in the classroom; and teacher involvement in evaluating the outcome of time spent together before, during, and after the coaches' visits.

Taken together, these two dimensions formed the basis for the scale used to measure each teacher's level of involvement with PIIC coaching at the treatment school. Actual data were gathered by the coach.

¹² Bean, R. (2004, spring). Promoting effective literacy instructions: The challenge for literacy coaches. *The California Reader*, 37(3), 58–63.

¹³ See Penn Literacy Network, <http://www.gse.upenn.edu/pln/>.

Teacher scores on the intensity of coaching measure were based on the following rubric:

- **Low:** All teachers at the treatment school received ongoing professional development and small group training sessions focused on PLN. This training figured into teacher classroom observations by administrators. By definition, because all teachers participated in this professional development program, everyone was at least coached at a low level. In addition to the formal training, the coach provided templates for the PLN strategies so they could be easily used in the classroom. In some cases, that was the extent of the coach's involvement with the teacher.
- **Medium:** These teachers specifically sought the coach's help with trying out what they saw in the professional development training. The coach went into classrooms and helped teachers incorporate new instructional strategies and PLN methods into their lessons. This typically involved a small amount of pre-planning as to what strategies fit best for a particular curriculum unit. Medium coaching involved a lot of modeling, co-teaching, and typically revisiting teachers periodically to see how they were doing, to share new ideas, and to plan additional visits.
- **High:** This involved responding to requests for help above and beyond an occasional visit to the classroom. Sometimes a teacher came to the coach asking for assistance on their own volition. Other times the coach noticed a teacher struggling and asked if it might be possible to work together on a regular basis. Other times an administrator asked the coach to provide ongoing assistance to a teacher having difficulty. In addition, there were some teachers who were coached to a high intensity because they liked the security of having help with planning, even though they may not have had a particular need. High intensity coaching sometimes followed schoolwide analysis of state data indicating where students were having particular difficulties. In these instances, the coach and the teacher may have spent considerable time planning and trying to pinpoint where and why students were having problems, and they worked together to create appropriate lesson plans. Sometimes the coach was in the classroom as much as once or twice a week and became effectively a co-teacher and co-facilitator.

Appendix B—Technical Notes

Statistical Tests

Although we had unit record data only on students in the treatment school, we were able to apply statistical tests to the grade-specific year-to-year differences in the percentage of students who were rated proficient or above on the PSSA test in reading (figures 2–7).

Assuming that each successive group of students entering a specific grade were unique (i.e., there were no students retained in grade) and that the two groups constituted random samples, we used a two-sample difference of proportions test. Thus, as shown in figure 2, the percentage for 3rd graders in the 2011–12 school year (82 percent) was significantly different from the percentage for 3rd graders in 2010–11 (41 percent). The significance level for this difference shown in the figure ($p < .05$) is the result of the test for the difference in proportions.¹⁴

We were unable to apply significance tests for the cohort data shown in figures 8–11 because we did not have unit record data for all students in the treatment *and* comparison schools that would have allowed for repeated measures' tests.

Data Sources

The data for this report are drawn from unit record annual reports of student performance on the Pennsylvania System of School Assessment in the treatment school and from aggregate data by grade for students in the comparison schools. These reports show, by grade, the number of students tested and the percentage who were rated advanced, proficient, basic, and below basic. Data are also available on performance described by student characteristics such as race/ethnicity, gender, and economic disadvantage.

As noted, unit record data on individual student performance in the *treatment* school were available to the researchers. These data were compared with the published data, and in a couple of instances, the percentage of students shown as proficient or advanced in the state reports varied by more than 5 percentage points from the same figures calculated from the unit record data. When such cases occurred, we substituted the unit record estimate for the published estimate and noted the substitution in the appendix tables.

¹⁴ STATA's immediate form of the test for a difference of proportions was used to estimate a two-tailed test of the difference in the percentage of students rated proficient or above.

Number of Students Scoring Proficient or Advanced on the PSSA, by Year, Grade, and Cohort—Treatment and Comparison Schools

Data for Figure 2: Number of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	30	23	32	33
Comparison A	36	33	34	48
Comparison B	34	31	45	30

Data for Figure 3: Number of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	29	14 ¹	25	27
Comparison A	47	38	36	38
Comparison B	43	37	33	45

¹ Constructed using student-level data file.

Data for Figure 4: Number of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	23	30 ¹	31 ¹	22
Comparison A	50	45	36	39
Comparison B	48	43	36	36

¹ Constructed using student-level data file.

Data for Figure 5: Number of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 3 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	13	11	22	15
Comparison A	16	22	23	36
Comparison B	24	25	31	27

Data for Figure 6: Number of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 4 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	17	13	13	17
Comparison A	29	20	21	28
Comparison B	30	31	26	32

Data for Figure 7: Number of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: Grade 5 2009–12

	School Year			
	2008–09	2009–10	2010–11	2011–12
Treatment	14	12	14	12
Comparison A	27	28	18	23
Comparison B	29	33	28	28

Data for Figure 8: Number of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5th through 8th grade cohort

	School Year			
	2008–09 Grade 5	2009–10 Grade 6	2010–11 Grade 7	2011–12 Grade 8
Treatment	23	24	28 ¹	28
Comparison B	48	48	44	42

¹ Constructed using student-level data file.

Data for Figure 9: Number of students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 6th through 8th grade cohort

	School Year		
	2008–09 Grade 6	2009–10 Grade 7	2010–11 Grade 8
Treatment	24	25	27
Comparison B	50	53	47

Data for Figure 10: Number of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 4th through 7th grade cohort

	School Year			
	2008–09 Grade 4	2009–10 Grade 5	2010–11 Grade 6	2011–12 Grade 7
Treatment	17	12	12	14
Comparison A	29	28	30	58
Comparison B	30	33	29	29

Data for Figure 11: Number of economically disadvantaged students who were rated proficient or advanced on PSSA reading in treatment and comparison schools: 5th through 8th grade cohort

	School Year			
	2008–09 Grade 5	2009–10 Grade 6	2010–11 Grade 7	2011–12 Grade 8
Treatment	14	14	17	15
Comparison B	29	35	25	23

Appendix C—Selected Studies of Instructional Coaching Outcomes

- Cornett, J., & Knight, J. (2008, July). *Research on coaching*. Instructional Coaching: Kansas Coaching Project. Lawrence, KS: Center for Research on Learning, University of Kansas. Retrieved November 6, 2012, from <http://instructionalcoach.org/research/tools/research-on-coaching>.
- Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Kohler, F. W., Crilley, K. M., Shearer, D. D., & Good, G. (1997). Effects of peer coaching on teacher and student outcomes. *The Journal of Educational Research*, 90(4), 240–260.
- Lockwood, J. R., McCombs, J. S., & Marsh, J. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis*, 32, 372–388.
- Marsh, J. A., McCombs, J. S., Lockwood, J. R., Martorell, F., Gershwin, D., & Naftel, S., et al. (2008). *Supporting literacy across the sunshine state: A study of Florida middle school reading coaches*. MG-762-CC. Santa Monica, CA: RAND Corporation.
- Marsh, J. A., McCombs, J. S., & Martorell, F. (2010, November). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, 24(6), 872–907. Retrieved November 6, 2012, from <http://epx.sagepub.com/content/24/6/872.full.pdf+html>.
- Medrich, E. A. (2008). *Coaching and student outcomes: Evidence from Philadelphia*. The Pennsylvania High School Coaching Initiative. Retrieved November 6, 2012, from <http://pacoaching.org>.
- Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Journal of Education*, 46, 532–566.
- Quint, J. (2011, July). *Professional development for teachers: What two rigorous studies tell us*. New York: MDRC.

- RAND Corporation. (2008). *Florida's middle school reading coaches: What do they do? Are they effective?* Research Brief. Santa Monica, CA: Author.
- Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education*, 17(1), 51–65.
- Sawchuk, S. (2011, August 18). Teacher-coaching boosts secondary scores, study finds. *Education Week, Teacher Beat*. Retrieved November 6, 2012, from http://blogs.edweek.org/edweek/teacherbeat/2011/08/teacher-coaching_boosts_scores.html.
- Veenman, S., Denessen, E., Gerrits, J., & Kenter, J. (2001). Evaluation of a coaching program for cooperating teachers. *Educational Studies*, 27(3), 317–340.
- Viadero, D. (2010). Coaching of teachers linked to stronger gains in reading. *Education Week*, 29(31), 6–7.

Appendix D—Characteristics of Treatment and Comparison Schools

	Treatment School	Comparison A	Comparison B
Grade Levels	K–8	K–8	1–8
Number of Students	234	476	329
Percent Economically Disadvantaged	54	65	71
Average Student Growth (PVAAS-AGI)*	-4.2	-5.7	-5.7

*AGI is a value based on the average growth across all grade levels and its relationship to the standard error so that comparison among schools is meaningful.

An AGI equal to zero means that the average achieving student in the school met the standard for PA academic growth. Average growth of less than zero means that, on average, students did not meet the standard for PA academic growth.