



DOCUMENTS REVIEWED:	“State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia.” “School Reform in Philadelphia: A Comparison of Student Achievement at Privately-Managed Schools with Student Achievement in Other District Schools.”
AUTHORS:	(“State Takeover...”): Brian Gill, Ron Zimmer, Jolley Christman and Suzanne Blanc (“School Reform...”): Paul Peterson
PUBLISHER/THINK TANKS:	(“State Takeover...”): RAND Corporation and Research For Action (RFA). (“School Reform...”): Program on Education Policy and Governance (PEPG) at Harvard University.
DOCUMENT RELEASE DATES:	(“State Takeover...”): Feb. 1, 2007 (“School Reform...”): April 10, 2007
REVIEW DATE:	May 7, 2007
REVIEWER:	Derek Briggs
E-MAIL ADDRESS:	Derek.Briggs@Colorado.EDU
PHONE NUMBER:	(303) 492-6320

Summary of Review

In 2002 the city of Philadelphia began a policy of restructuring its lowest-achieving elementary and middle schools. Eighty-six schools were included. Restructuring can take on a wide variety of forms, but in Philadelphia the most prominent approaches shifted school management to either the district or one of several private providers. In 2007, after four years of this policy, two research reports were issued, one by RAND in collaboration with Research For Action (RAND-RFA) and one by the Program on Education Policy and Governance (PEPG). Both reports examined whether any positive effects on the math and reading achievement of students could be attributed to privately managed schools, district-managed schools, or neither. According to the RAND-RFA report, private management has had no cumulative effect on math or reading achievement, while district management has had a positive effect on math achievement but no effect on reading. According to the PEPG report, private management has had a positive effect on the percentage of students reaching “Basic” levels of performance in math and reading, while district management has generally had no effect. The different findings from the two reports can largely be explained by the fact that PEPG did not have the same access to data as did RAND-RFA. PEPG also analyzed data using a different methodological approach than did RAND-RFA, due in large part to the data limitations. This review identifies and describes methodological weaknesses in the report from RAND-RFA as well as in the PEPG report. Overall, while the RAND-RFA study appears to better capture the overall effects of Philadelphia’s reform than does the PEPG study, it does not differentiate effects between the elementary and middle school grades. Further analysis and research is needed before drawing any definitive conclusions.

Review

I. INTRODUCTION

When an American public school repeatedly fails to demonstrate progress in meeting the academic performance standards established by a statewide accountability system (i.e., it fails to make “adequate yearly progress”), what should be done? According to the provisions of the No Child Left Behind (NCLB) Act:

If a school fails to make adequate yearly progress [in meeting proficiency standards] for a fifth year, the school district must initiate plans for *restructuring* the school. This may include reopening the school as a charter school, replacing all or most of the school staff, or turning over school operations either to the state or to a private company with a demonstrated record of effectiveness.¹

As of the summer of 2008, states will begin identifying schools that have failed to make adequate yearly progress under NCLB for five consecutive years. The number of schools identified as failing is likely to grow with each subsequent year as the difficulty of meeting the requirements for adequate yearly progress increases.² Assuming no dramatic changes are made to NCLB over the next few years, states and school districts will be faced with a difficult decision: what approach to restructuring schools is most likely to lead to improvements in academic achievement?

In 2002, just as the NCLB Act was going into effect, the city of Philadelphia was already in the process of restructuring 86 chronically low-achieving elementary and middle schools located in the School District

of Philadelphia (SDP). The schools were to be restructured according to what became known as the “diverse provider model.” Under this model, school restructuring was to occur in four different ways:

1. Forty-five (45) schools were to receive private management from one of seven for-profit, nonprofit, and university providers (“private management”);
2. Twenty-one (21) schools were to be managed by the SDP (“district management”);
3. Sixteen (16) schools were to continue to manage themselves (known as the “sweet-sixteen” schools); and
4. Four (4) schools were to be transformed into charter schools.

Beyond these differences in school management, all of these 86 schools were to be given additional financial resources. Now, four years later, one can compare changes in student achievement among schools under these different conditions relative to one another, and relative to other schools in Philadelphia that were not restructured. Two different reports have performed such comparisons but have seemingly arrived at opposite conclusions about the efficacy of private and district management.

The first report, “State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia,” was written by Brian Gill, Ron Zimmer, Jolley Christman and Suzanne Blanc of the RAND Corporation and the organization Research For Action (RFA).³ The second report, “School Reform in Philadelphia: A Comparison of Student Achievement at Privately-Managed Schools with Student Achievement in Other District Schools” was

written by Paul Peterson, a faculty member at Harvard University's Kennedy School of Government and director of the Program on Education Policy and Governance (PEPG).⁴ The PEPG published this second report. This review explores how the differing conclusions reached by the RAND-RFA and PEPG reports might be reconciled.

II. FINDINGS AND CONCLUSIONS OF THE TWO REPORTS

The RAND-RFA Report

The RAND-RFA report consists of two different analyses: a “districtwide analysis” and a “diverse-providers analysis.” In the districtwide analysis, changes in aggregate student achievement for schools in the SDP from 2001 (pre-intervention) to 2006 (four years after the restructuring intervention began) were compared with changes in similarly low-achieving schools across the state of Pennsylvania. This analysis was conducted to test the hypothesis that the diverse-provider model—by promoting internal competition—might spur improvements in low-achieving schools across the SDP, whether the schools had been restructured or not. In the diverse-providers analysis, the authors compare changes in student achievement for schools within the SDP, as a function of whether and how the school was restructured as of 2002. Of primary interest in this second analysis was whether privately managed schools, district managed schools, or the sweet sixteen schools had gains larger than those found in schools that were not restructured or given any additional resources. (Students in the four schools that were being restructured into charter schools were excluded from all analyses.) Of secondary interest was whether the relative gains for privately managed schools varied

among the seven for-profit, nonprofit and university management providers.

For its district-wide analysis, conducted with school-level data, the RAND-RFA team found that the percentage of elementary and middle-school students classified as “proficient” based on their performance in reading and math on the Pennsylvania State System of Assessment (PSSA) across the SDP had increased substantially (by about 10 percentage points in reading, and by a little over 20 points in math) in the years since the state takeover. When these gains were compared with the gains of low-achieving schools in other parts of the state, statistically significant differences that favored SDP schools were found for grade 5 and grade 8 students after three years of the intervention. By the fourth year of the intervention, however, an SDP advantage could only be found for grade 8 reading scores, which were five percentage points higher in SDP schools. On all other available school measures of student proficiency, the four-year proficiency gains of the SDP schools from pre-intervention baseline scores were “indistinguishable” from the gains of state-wide comparison schools.⁵

While the SDP test scores during this period of time may have been affected by the restructuring policy, those scores may also have been affected by a number of other factors, including several additional districtwide initiatives, enacted from 2002 through 2005.⁶ While many of these initiatives were applied to both restructured and non-restructured schools, some were specific only to the non-restructured schools. In this regard, and relevant to the discussion below, it is important to note that the key control group in both the RAND-RFA and PEPG reports—non-restructured SDP schools—does not represent schools in

which no other interventions were taking place.

The RAND-RFA team used student-level data for its diverse-provider analysis. Team members found that the cumulative four-year achievement gains among privately managed schools and among sweet sixteen schools were no different than the gains among non-restructured schools in the SDP. In contrast, a positive and moderately sized effect⁷ on math achievement was found for schools that had been restructured under district management. When privately managed schools were disaggregated by provider, few statistically significant effects were found—with the exception of large negative effects in both reading and math for schools managed by Temple University, and in math for schools managed by the Victory organization. The RAND-RFA report concludes: “with four years of experience, we find no evidence of differential academic benefits that would support the additional expenditures on private managers.”⁸

The PEPG Report

The PEPG report was written in response to the findings of the RAND-RFA report. In fact, much of the PEPG’s report consists of criticisms of the findings of the RAND-RFA report and of the methods used by the RAND-RFA authors to arrive at these findings. But the PEPG report also included an original analysis. Professor Peterson uses school-level data to compare the changes in the percentage of students classified at different performance levels on the PSSA math and reading tests. He compares data from two cohorts of students enrolled in restructured schools that were privately managed or district managed to the change observed for all other SDP schools that had not been restructured.⁹ The first cohort consisted of students who were in grade 5 in 2002 and

grade 8 in 2005. The second cohort consisted of students who were in grade 5 in 2003 and grade 8 in 2006.

Peterson finds that for both cohorts the average increases in the percentage of students scoring at or above the “Basic” level on the PSSA in both reading and math for privately managed schools was between 4 and 13 percent larger than those found for schools that had not been restructured or had been district managed. Similar effects were not found for privately managed schools when the outcome of interest focused on increases in the percentage of students scoring at or above the “Proficient” level. That is, his positive findings were limited to lower-scoring students. In neither case did Peterson find the positive effects for publicly managed schools that were indicated by the RAND-RFA report.

III. THE REPORTS’ USE OF RESEARCH LITERATURE

The RAND-RFA report draws heavily from the work of Paul Hill¹⁰ to establish the theoretical rationale for why restructuring schools through the use of a diverse-provider model would be expected to increase student achievement. In theory, opening the management of schools to many possible providers is expected to establish a “competitive school marketplace in which districts manage a varied portfolio of schools, providers have wide rein to innovate, and both are held accountable for student outcomes by strong contracts and through the availability of meaningful choices for schools and parents.”¹¹ However, as the RAND-RFA report makes clear, the diverse provider model implemented in Philadelphia diverged from Hill’s ideal, primarily in that students and parents were not given the option to choose their own schools and, in addition, constraints were

placed on the management options of external providers.

Neither report provides much of a conceptual framework for the theoretical efficacy of privately managed or district managed schools. According to the RAND-RFA report, empirical findings as to the efficacy of privately managed schools on student achievement have been mixed. The RAND-RFA team cites a prior large-scale evaluation of the Edison schools¹² in which it was found that while Edison-operated schools appear to have no effects on student achievement in the first three years of implementation, positive effects emerge in years four and five.

IV. REVIEW OF THE REPORTS' METHODOLOGIES

Professor Peterson is justifiably critical of the approach taken in the first, districtwide analysis conducted by the RAND-RFA team. In that analysis, increases in the percentage of students scoring at the “Proficient” level of the PSSA among schools in the SDP were compared to schools in other Pennsylvania school districts. The RAND-RFA authors had restricted the samples to only those schools in the lowest quartile of the statewide achievement distribution. This appears problematic for two reasons: First, it has the effect of eliminating between 17 and 35 percent of the SDP schools from the comparison (those above the lowest statewide quartile). Second, no descriptive statistics are provided by the RAND-RFA authors to make the case that the two groups are comparable in terms of readily available demographic characteristics. The diverse provider model may have had districtwide effects on the achievement of students in Philadelphia schools, but there are strong reasons to question the validity of the effects estimated in the RAND-RFA analysis.

In order to explore why the two reports arrived at different conclusions about the effectiveness of privately managed and district managed schools, one must first understand the differences in the data analyzed in each report. Figure 1 illustrates the general structure of the data made available to the RAND-RFA researchers by the SDP. Each cell in the figure represents the math and reading test scores for a cross-section of students in over 200 schools in the SDP for a particular grade and year. The letter in each cell indicates the type of test that was administered to students. For example, fifth grade students took PSSA tests in math and reading in 2002, but that cohort took the Terra Nova tests in math and reading as sixth grade students in 2003.¹³ Test scores in 2001 can be considered baseline pre-intervention observations; test scores from 2002 forward represent post-intervention observations. Each of the 12 diagonal lines in Figure 1 represents a distinct longitudinal cohort of students. All cohorts that included test scores in math and reading for at least two points in time were included in the RAND-RFA analysis (10 out of 12 cohorts). In contrast, the four rose-shaded cells within two of these 10 cohorts represent the subset of data that was the basis for the findings in the PEPG report. These were the only data publicly available, and they were only available at the school level, not the student level.

There were two key decisions made by the RAND-RFA team in how they chose to analyze the data represented in Figure 1. First, the authors decided to specify a single statistical model—known as a fixed effects model—that incorporated information about test score changes from the 10 different student cohorts over the 2001 to 2006 time period. Second, the RAND-RFA team standardized test scores in each grade and year relative to all schools in the SDP as a means of establishing a common score scale out of

Figure 1. Structure of the Data Analyzed in RAND-RFA and PEPG Reports

GRADE	2001	2002	2003	2004	2005	2006
2			F	F	F	P
3	S	S	F	F	F	P
4	S	S	F	F	F	P
5	P	P	P, T	P, T	P, T	P
6			T	F	F	P
7	S	S	T	F	F	P
8	P	P	P, T	P, F	P, F	P

Note: P = PSSA Tests, S = Stanford 9 Tests, T = Terra Nova Tests

the three different tests that were administered.¹⁴

In his report, Peterson criticizes the RAND-RFA methodological approach for the following reasons:

1. Insufficient context is provided to help the reader evaluate the use of their fixed effects model.
2. Combining the results from different tests, some of which are high-stakes (PSSA) and some of which as low-stakes (Terra-Nova and Stanford 9), may lead to the underestimation of the effects of privately managed schools because students at these schools might take low-stakes tests less seriously.
3. The RAND-RFA analysis does not constitute a quasi-experimental analysis because “Instead of comparing gains during the pre-treatment with gains post-treatment, the researchers simply compared levels of achievement... As a result, their study is seriously at risk of having under-estimated the impact of the privately-managed schools...”¹⁵

Professor Peterson makes the first of these three criticisms only implicitly, but it is en-

tirely on target. There are many different statistical models the RAND-RFA team could have specified to estimate the effects of different approaches to school reconstruction on student achievement, and the rationale for the one they have chosen to use is never entirely clear.¹⁶ Even if one takes as a given the objective of applying a single model to the full set of longitudinal cohorts represented in Figure 1, one could just as readily specify what is known as a random effects model instead of a fixed effects model. Of course, different approaches have different advantages and disadvantages, but the RAND-RFA report never explains why the authors have rejected other modeling alternatives in favor of their fixed effects model specification. A technical appendix with such information—with detail on any sensitivity analyses that were conducted—would have considerably strengthened the warrant for the report’s conclusions.

Peterson’s second criticism is less credible. To begin with, standardized tests in an accountability context are primarily high stakes for teachers and schools—not necessarily the students who take them. As a threat to the validity of the RAND-RFA findings, problems induced through the use of different tests from year to year will mostly be mitigated by the fact that students

in all schools under comparison also took these different tests. Hence, even if SDP students change their effort when moving from high- to low-stakes tests, there is no reason to suspect that students in privately managed schools took these tests less seriously than students in any of the other SDP schools. The estimated effect of private management would only be underestimated if this differential effort were the case, and Peterson provides no evidence to support this supposition. It would be just as likely that the estimated effect of private management was overestimated because students in the non-restructured SDP schools took the tests less seriously than their privately managed counterparts.¹⁷

In Peterson’s third criticism he claims that the RAND-RFA team has—notwithstanding the authors’ claims—failed to conduct what is known as a “quasi-experiment” to estimate the effects of private and district management. After providing his own definition for what constitutes a quasi-experiment, Peterson writes “Had the RAND-RFA conducted such a quasi-experiment, its results would have considerable credibility. But the RAND-RFA did not conduct a quasi-experiment. Instead of comparing gains during pre-treatment with gains post-treatment, the researchers simply compared levels of achievement.”¹⁸ While it is entirely reasonable for Peterson to raise questions about the warrant for the conclusions the RAND-RFA team draws from their statistical analysis, he is wrong to claim that they have not conducted a quasi-experiment. In their book, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Shadish, Cook and Campbell define an experiment as “a study in which an intervention is deliberately introduced to observe its effects” and a quasi-experiment as “an experiment in which units [i.e., schools] are not assigned to conditions randomly.”¹⁹

Clearly then, the RAND-RFP study represents a quasi-experiment—SDP schools and students have been non-randomly assigned to one of four interventions or to the control. Hence any subsequent analysis of this data represents a quasi-experimental analysis.

However, not all quasi-experimental analyses are equally strong as a basis for subsequent causal inferences. Peterson favors a particular methodological approach, and because this is not the approach taken by the RAND-RFA team he implies that the study’s fundamental conclusions about the ineffectiveness of private management (and effectiveness of district management) are invalid. To illustrate the distinction between the approach Peterson prefers and the approach that the RAND-RFA team has taken, imagine that we have annual student test score data from two types of schools for the years 2000, 2001 and 2002. The “control” schools maintain the same management from 2000 to 2002; the “treatment” schools switch to private management as of 2002 (that is, for the final year). The cells in Table 1 indicate the test score averages that would be available in each year for the treatment and control schools.

Table 1. Illustration of Peterson’s Critique of RAND-RFA Diverse Provider Analyses

	2000	2001	2002
Treatment Schools	\bar{X}_{t00}	\bar{X}_{t01}	\bar{Y}_{t02}
Control Schools	\bar{X}_{c00}	\bar{X}_{c01}	\bar{Y}_{c02}

Under Peterson’s methodological approach, the effect of the treatment would be estimated as

$$\begin{aligned} &[(\bar{Y}_{t02} - \bar{X}_{t01}) - (\bar{X}_{t01} - \bar{X}_{t00})] - \\ &[(\bar{Y}_{c02} - \bar{X}_{c01}) - (\bar{X}_{c01} - \bar{X}_{c00})] \end{aligned}$$

This mathematical expression communicates the idea that for both treatment and control schools, pre-treatment score gains are being compared to post-treatment score gains. Under the RAND-RFA approach (which would not use the available data for 2000, primarily because it would reduce their available sample size), the effect of the treatment would be estimated as

$$(\bar{Y}_{t02} - \bar{X}_{t01}) - (\bar{Y}_{c02} - \bar{X}_{c01})$$

That is, for both treatment and control schools, pre-treatment score levels (as opposed to gains) are being compared to post-treatment score levels. The two approaches will, in fact, produce the same results if the pre-treatment score gains are the same for both treatment and control schools—i.e., when

$$(\bar{X}_{t01} - \bar{X}_{t00}) = (\bar{X}_{c01} - \bar{X}_{c00})$$

The RAND-RFA team is able to present evidence that this equality holds for their data, which appears to refute this aspect of Peterson’s criticism.²⁰

Nonetheless, there are other assumptions of the RAND-RFA modeling approach that could be called into question. While some of these assumptions are briefly discussed in their report, few are given sufficient scrutiny. For example, the fixed effects model assumes that all unobserved variables that might confound the estimated effect of private management (e.g., student motivation and socioeconomic status) stay constant over time, such that when pre- to post-intervention scores are computed for each school, the influence of these potentially confounding variables will wash out. This

seems somewhat implausible. It is an assumption that will be violated, for example, if students at privately managed schools become increasingly motivated to do well academically with each passing year. Yet if the intervention is working, one might expect to observe exactly those types of changes. This is the sort of issue that suggests a possible tension between the match of the RAND-RFA team’s statistical model to the actual situation in the schools. The authors are not entirely successful in reassuring the reader that these sorts of tensions do not compromise the validity of their findings.

While the methods employed in the RAND-RFA analysis place it on one end of a modeling complexity continuum, the methods employed by the PEPG analysis place it on the opposite end. On the one hand, it is a strength of Peterson’s analysis that his approach for estimating the effects of private and district management is easy to follow. On the other hand, the simplicity of Peterson’s approach is also a weakness, because potential threats to the validity of his inferences are especially difficult to rule out. There are a number of potentially serious problems with the methodological approach in the PEPG report, and they are summarized below.

1. The baseline year for PEPG’s comparisons of grade 5 to grade 8 growth in each of the two cohorts (2002, 2003) does not precede the restructuring interventions. This means that any initial effects of the reform are not captured by the PEPG analysis. Moreover, because the PEPG baseline year is different than the baseline year (2001) used in the RAND-RFA analysis, the different effects estimated by the two reports are hard to compare.

Table 2. Illustration of Subset of Private and District Providers Included in the PEPG Analysis using 2006-07 Data.

Provider	Range of Grades in Restructured Philadelphia Schools										
	K-5	K-6	K-7	K-8	5-8	6-8	7-8	8-9	7-12	9-12	TOTAL
Edison		4		8	4	2	2				20
Foundations		4		1			2			1	8
Temple				5				1			6
University of Penn		1		2							3
Victory		2			1				1		4
District	2	4	2	9			1		1	2	21
TOTAL	2	15	2	25	5	2	5	1	2	3	62
% of Column Total	3.2%	24.2%	3.2%	40.3%	8.1%	3.2%	8.1%	1.6%	3.2%	4.8%	100%

2. The PEPG report restricts the number of schools included in the privately managed and district managed categories to those that include both grade 5 and grade 8. This is illustrated in Table 2. The two rose-shaded (K-8 and 5-8) columns indicate the subset of schools that would be included in PEPG’s analysis based on 2006-2007 data. However, Professor Peterson does not appear to impose the same restriction for the non-restructured district schools being used as his comparison group. As a result, the average gains for students who remain in any given privately managed school over the four-year period are being compared to the average gains for two distinct groups of students: (a) those who stay in the same non-restructured school from grade 5 to grade 8, and (b) those who switch from an elementary to a middle school. If, for example, switching schools has a negative association with student achievement, this would lead to the overestimation of the effects Peterson has found for both privately managed and district-managed schools.

3. Based only on the school-level summaries of student achievement data available to Peterson, one has no way of knowing how many of the students represented in a given school’s grade 5 test scores are subsequently represented in the grade 8 test scores for the same school. If the students with the highest test scores are more likely to stay in a given type of restructured school while the students with the lowest scores are more likely to leave, any effect of restructuring would be overestimated. If the converse were true, any effect would be underestimated.

4. As noted above in (2), Peterson’s available sample of schools is restricted because SDP schools vary in terms of the grade span they include. His approach for comparing test-score change over time only works for schools that have students in both grade 5 and grade 8. Because he is only including data for about 40 percent of all privately managed schools, and between 20 and 40 percent of all district-managed schools in his two cohort analyses, it is un-

clear whether his estimated effects generalize to the unrepresented schools, even if these effects are accurate.²¹

V. REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS

Though the findings from these two reports appear contradictory, it is possible that both are correct. There are at least two reasons this might be so. One possibility Peterson puts forward is that private management may have a negative effect in elementary grades, but a positive effect in middle grades. When an overall effect across grades 3 through 8 is estimated in the RAND-RFA report, the negative early grade effect may cancel the positive middle grade effect. Neither report, however, provides a cogent explanation for why this would be expected.

A second possibility concerns the fact that the PEPG analysis uses only categorical test score information (i.e., the percentage of students in a school that fall into the “Basic” and “Proficient” performance levels on the PSSA). In fact, Peterson only found positive effects for privately managed schools when the outcome of interest was the change in the percentage of students at the “Basic” (but not the higher, “Proficient”) level. This outcome (improvement for students scoring at the Basic level) is a function of movement among students at the lowest end of the PSSA score scale. By contrast, the estimated effects summarized in the RAND-RFA analysis represent the average score gain across the full PSSA, Stanford 9 or Terra Nova standardized score scale for any given grade and year.²² The take-home point is that while the PEPG analysis has found positive effects for private management, these effects are based upon both a very specific subgroup of schools (those

containing grades 5 through 8), and a very specific subgroup of students (those who have performed the worst on the PSSA test). The RAND-RFA analysis, on the other hand, found no overall effect for private management on the basis of all the SDP schools and all the SDP students.

To the extent that RAND-RFA researchers still have access to the SDP data, it should be relatively straightforward to reconcile these differences empirically. If Peterson’s school-level analysis were to be done with student-level test scores, using only students attending the same school in grades 5 and 8, would the estimated effects for privately managed and district managed schools more closely resemble those found by Peterson, or would they resemble the overall effects presented in the RAND-RFA analysis? This would be important to find out, because if private and district management have different effects on the achievement of elementary and middle-school students (as the PEPG analysis suggests), this finding would have clear policy ramifications.

A clear strength of both reports is that they are each well-written, cognizant of the limitations of their analyses, and relatively cautious in generalizing their findings. A weakness of both reports—particularly the RAND-RFA report—is a lack of important descriptive statistics. A prime example of this is the missing table or figure that would detail the movement of students in and out of the schools in the SDP during the time period being analyzed. Neither report gives the reader any sense of the variability of test score gains within schools or across students.

Of the two reports, the RAND-RFA report has the advantage of better and more comprehensive data during the 2001 to 2006 time period of interest in Philadelphia. Be-

cause of this, the conclusions from this report about the efficacy of private and district management on student achievement are likely to carry more weight than those from the PEPG report, and this appears justifiable. There are threats to the validity of both reports, but because the PEPG report is only able to analyze a subset of the RAND-RFA data at the school rather than at the student level, there are more reasons to doubt the PEPG findings.

VI. USEFULNESS OF THE REPORTS FOR GUIDANCE OF POLICY AND PRACTICE

The results from the RAND-RFA report suggest that, on average, schools that were restructured through district management in Philadelphia were more effective at increasing math achievement among students than comparable schools that were restructured with private managers. Also, none of the restructuring interventions had lasting and appreciable effects on reading achievement. In contrast, the PEPG report suggests that privately managed schools containing the grades 5 and 8 are in fact having a positive effect on the reading and math achievement of students in the lowest part of the PSSA test score distribution.

It is recommended that neither of these reports be used as the primary basis for policy

decisions until subsequent analyses have been performed to provide a more nuanced view of the effects of the restructuring interventions at different grade levels, and at different locations along the PSSA test score distribution.

Nevertheless, both of these reports appear to be having an immediate impact on policy and practice. The RAND-RFA report is currently being used by Philadelphia's Accountability Review Council as the justification for recommending the firing of underperforming private managers and as the basis for identifying the "schooling and organizational conditions that contributed to the success of the District restructured schools."²³ Meanwhile the findings from the PEPG report have been reported in the *Wall Street Journal* and other media outlets.²⁴

Given NCLB's nod toward school restructuring, the outcomes in Philadelphia have national import. These two studies, read together, do contribute to our understanding of Philadelphia's experience. Yet, although the RAND-RFA study sheds more light on that experience than does the PEPG study, neither study offers a complete picture, and more research is needed before drawing any definitive conclusions.

NOTES & REFERENCES

- ¹ U.S. Department of Education. (2003) *No Child Left Behind: A Parents Guide*. Washington, D.C., p. 9.
- ² Linn, R. L. (2003). "Accountability: Responsibility and Responsible Expectations." *Educational Researcher*. Vol. 32, No. 7, 3-13.
- ³ Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ⁴ Peterson, P. (2007) School reform in Philadelphia: A comparison of student achievement at privately managed schools with student achievement in other district schools. Kennedy School of Government, Harvard University, Program on Education Policy and Governance.
- ⁵ See p. xii in Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ⁶ See Table 2.2, p. 10 in Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ⁷ The actual magnitude of the effect, expressed as a proportion of the variability typical of math test scores (i.e., the standard deviation), was 0.2.
- ⁸ See p. 41 of Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ⁹ Peterson does not distinguish the sweet 16 schools separately in his analysis. It appears that even though they received additional resources, they are grouped with all other non-restructured schools.
- ¹⁰ Hill, P. T. *Put learning first: A portfolio approach to public schools*. Washington, D.C.: Progressive Policy Institute, February 2006.
- ¹¹ See pp. 8-9 of Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ¹² See Gill, B., Hamilton, L., Lockwood, J. R., Marsh, J., Zimmer, R. Hill, D., and Pribesh S. *Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools*, Santa Monica, Calif.: RAND Corporation, MG-351-EDU, 2005. <http://www.rand.org/pubs/monographs/MG351/> Note that this evaluation was also conducted by RAND and had the same lead author (Brian Gill) as the current RAND-RFA report.
- ¹³ Note that for the second grades and sixth grades, no tests were administered in 2002 and 2003. In the fifth and eighth grades two different tests were administered from 2003 through 2005.
- ¹⁴ Because SDP students did not take the same tests across grades, there is no criterion-referenced measure available to assess growth in scores over time. Test scores were standardized by the RAND-RFA team to get around this problem. For example, in 2001 test scores were available for grade 4 students in the SDP that took the Stanford 9 math test. The RAND-RFA team took the score of each student in this sample, subtracted the average Stanford 9 math score for the SDP as a whole, and then divided by the standard deviation of the test scores. After doing this, all students had standardized scores that ranged between about 3 standard deviations above and below the district average. By the next year (2002), these same students were now in grade 5, but were administered the PSSA math test, which has a

-
- different score scale than the Stanford 9. Accordingly, the RAND-RFA team took the score of each student in this sample, subtracted the average PSSA math score for the SDP as a whole, and then divided by the standard deviation of the test scores. Because of this, all students once again had standardized scores that ranged between about 3 standard deviations above and below the district average. The upshot of this approach is that changes in test scores from grade 4 to grade 5 can only be given a relative interpretation in the RAND-RFA analysis: Has a student's achievement improved relative to his or her district peers?
- ¹⁵ See pp. 14-17 of Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ¹⁶ For example, a more transparent approach might have been to specify separate statistical models for each of the four longitudinal cohorts of students that were tested in the baseline year of 2001.
- ¹⁷ An interesting supposition that might be more plausible is that some of the math and reading curricula associated with schools under different forms of management may be differentially aligned to the three tests being combined onto a common scale. If, for example, Edison's 'Success for All' curriculum is more aligned with the PSSA tests than the Terra Nova tests, while the curricula used by non-restructured schools are more aligned to the Terra Nova tests, this might bias any estimated effect for an Edison school when test score changes are based on a transition from PSSA to Terra Nova.
- ¹⁸ See pp. 14-15 in Peterson, P. (2007) School reform in Philadelphia: A comparison of student achievement at privately managed schools with student achievement in other district schools. Kennedy School of Government, Harvard University, Program on Education Policy and Governance.
- ¹⁹ See p. 12 in Shadish, W., Cook, T., and Campbell, D. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*.
- ²⁰ See Table 4.1 on p. 31 of Gill, B., Zimmer, R., Christman, J. & Blanc, S. (2007) *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: The RAND Corporation.
- ²¹ Peterson shows that the racial/ethnic mix and free and reduced lunch status of students in restructured schools included in the analysis is comparable to that of the restructured schools excluded in the analysis. However, there is little variability in these characteristics among the restructured schools, so this comes as little surprise. What would be more compelling would be a comparison of the average grade 5 or grade 8 test scores for the included and excluded schools.
- ²² It is entirely possible that the difference between the RAND-RFA findings and the PEPG findings might be captured by the inclusion of a series of dummy variables in the fixed effects model that represent the interaction between a given restructuring intervention and a student's location in the lowest achievement quartile in a prior year.
- ²³ Accountability Review Council. (2007) Report to the School Reform Commission: The Status of 2005-2006 Academic Performance in the School District of Philadelphia. February, 2007.
- ²⁴ See Peterson, P. (2007) The Philadelphia Story. *The Wall Street Journal*. February 23, 2007. Retrieved April 25 from http://www.ksg.harvard.edu/pepg/pr/WSJOPED_20070222.htm
-

The Think Tank Review Project is made possible by funding from the Great Lakes Center for Education Research and Practice.