# Impact of For-Profit and Non-Profit Management on Student Achievement:

# The Philadelphia Experiment

**Paul E. Peterson and Matthew M. Chingos**

**PEPG 07-07**

**Prepared under the auspices**

**of the**

**Program on Education Policy and Governance**

**Kennedy School of Government**

**Harvard University**

**November 1, 2007**

**Paul E. Peterson is the Henry Lee Shattuck Professor of Government and Director of the Program on Education Policy and Governance in the Kennedy School of Government. He is a Senior Fellow at the Hoover Institution, Stanford University**

**Matthew Chingos is a Research Fellow in the Program on Education Policy and Governance.**

**Program on Education Policy and Governance**
**Kennedy School of Government**
**79 Kennedy Street, Taubman 306**
**Cambridge, MA 02138**
**Tel: 617-495-7976/8312**
**Fax: 617-496-4428**
**www.ksg.harvard.edu/pepg/**

**Impact of For-Profit and Non-Profit Management on Student Achievement:**

**The Philadelphia Experiment**

***Paul E. Peterson and Matthew M. Chingos***

## Executive Summary

The School District of Philadelphia, in the summer of 2002, at the request of the State of Pennsylvania, asked for-profit and non-profit managers to participate in a substantial restructuring of its lowest-performing schools under the overall direction of the Philadelphia School Reform Commission (SRC). Thirty elementary and middle schools were contracted out to for-profit management organizations; 16 schools were contracted out to non-profit organizations. Another 21 schools were assigned to be restructured by a newly created Office of Restructured Schools (ORS), a special office within the school district itself. Sixteen schools (commonly called the "sweet sixteen") were simply given additional fiscal resources.

Using individual student test-score data made available by the School District of Philadelphia, we estimated the impact of for-profit and non-profit management on student achievement by tracking the performance of students in math and reading from 2001 to 2006. The first two years (2001 and 2002) provide us with information about student performance prior to the management intervention, while the subsequent four years (2003, 2004, 2005, and 2006) provide information about student performance after the interventions had begun. The measures of test performance are taken from three tests: the Terra Nova, the Stanford 9, and the Pennsylvania State System of Assessment (PSSA), the test Pennsylvania administers to comply with the accountability requirements in the federal No Child Left Behind Act (NCLB).

Of the 30 schools included in the study that were under for-profit management, 20 were managed by Edison Schools, five by Victory Schools, and five by Chancellor Beacon Academies. Of the 16 schools included in the study that were managed by non-profits, five were managed by Foundations, three by the University of Pennsylvania, five by Temple University, and three by Universal Companies. We compare the performance of the privately managed schools to that of 71 schools that remained under regular school district management whose students performed below the district median but were not included in the "sweet sixteen" group of schools.

After three and four years, the average student at schools managed by for-profit firms learned more in math than would be expected had the schools remained under district management. The impact of for-profit management on average math performance was large—approximately 26 percent and 29 percent of a standard deviation, respectively, or roughly two-thirds of a year of additional learning over a four-year period of time. However, the positive impact of for-profit management on average reading gains after three and four years was only 7 to 8 percent of a standard deviation, an impact that was not statistically significant.

For non-profits, we found mainly negative impacts on student performance in both math and reading. In the third year the size of the negative effect was 15 percent of a standard deviation in math and 9 percent of a standard deviation in reading. In year four the negative impact in math was 16 percent of a standard deviation, but the reading impact was positive, at 4 percent of a standard deviation. However, none of those impacts was statistically significant. (The estimated impacts of the restructured and sweet sixteen schools on student performance were generally small and always statistically insignificant in both subjects in all years.)

Finally, schools under for-profit management out-performed those under non-profit management by sizeable margins. After four years, students in schools under for-profit management learned, on average, over a year's worth of math more than those in schools under non-profit management. That difference in impact was statistically significant.

The average impact of attending a school under for-profit or non-profit management (relative to that school's remaining under district management) was identified using a quasi-experimental research design known as "difference-in-differences" analysis. The treatment group consisted of schools managed by each type of provider, and the control group included the 71 schools under regular district management at which students were performing below the district median. To estimate treatment effects, we controlled for student and school fixed effects and calculated the difference between test-score performance gains in the years before and after the intervention began at treatment and control schools. So, for example, if test-score gains at the treated schools were 20 percent of a standard deviation higher than gains before the intervention, while gains over that period of time at the control schools were only 15 percent of a standard deviation higher, the estimated effect of for-profit management would be the difference between these, or 5 percent of a standard deviation.

A prior investigation conducted by RAND-Research for Action (RAND-RFA) reported no impact of private management on student test-score performance in either reading or mathematics. Our study differs from the RAND-RFA study in that our study 1) estimates impacts separately for schools under for-profit and non-profit management; 2) compares the schools under private management with just the other low-performing schools in Philadelphia; 3) compares student gains (not just their levels of performance) before and after treatment; 4) includes students as long as we can calculate two gain scores for them, even if they do not remain in the same school; and 5) controls for student movement across schools.

**Impact of For-Profit and Non-Profit Management on Student Achievement: The Philadelphia Experiment**

***Paul E. Peterson and Matthew M. Chingos***

No Child Left Behind (NCLB) asks that states "restructure" schools that fail for six years running to make Adequate Yearly Progress (AYP) toward full proficiency on the part of all students by the year 2014.[1] The law provides a number of restructuring options, including the shift of the school's management to a private entity, either a for-profit firm or a non-profit organization. Only a few school districts nationwide have sought help from either for-profit or non-profit organizations in the management of low-performing schools (Mead 2007). The School District of Philadelphia, in the summer of 2002, at the request of the State of Pennsylvania, asked both types of entities to participate in a substantial restructuring of many of its lowest-performing schools under the overall direction of the Philadelphia School Reform Commission (SRC). Thirty elementary and middle schools were contracted out to for-profit management organizations,[2] 16 schools were contracted out to non-profit organizations, and 21 schools were

[1] We wish to thank the School District of Philadelphia's Office of Accountability, Assessment, and Intervention for providing the information on student performance. We also express our appreciation to Terry Moe, Eric Hanushek , Steve Rivkin, and Martin West for their methodological and substantive comments on this paper. Antonio Wendland and Ron Berry provided staff assistance. The research for this paper is supported in part by the National Research and Development Center on School Choice, Competition, and Achievement, which is funded by the Department of Education's Institute of Education Sciences (R305A040043) as well as by grants from the John M. Olin Foundation, the Lynde and Harry Bradley Foundation, and Edison Schools. Grantors were given the opportunity to review the manuscript only for the purpose of identifying factual errors. The findings and interpretation are the sole responsibility of the authors.

[2] A high school was contracted out to a for-profit provider in 2004-05.

assigned to be restructured by a newly created Office of Restructured Schools (ORS), a special office within the school district itself (Herold and Riffer 2005).

The management of the restructured schools was to take place within the terms of collective bargaining agreements with employee unions. Other district policies were to remain in effect as well. The managers' tasks were greatly complicated by the SRC decision to allow teachers at the schools to transfer to another school in the district, if they so desired. Finally, the for-profit entities were asked to manage the schools that had the very lowest-performing students, and the students attending schools assigned to the non-profits had only marginally higher test-score performance than students in the schools assigned to the for-profits.

The policy intervention in Philadelphia raises questions of general interest: Will students at schools assigned to for-profit or non-profit managers learn more than would be expected had those schools remained under traditional school district management? Is for-profit management more or less effective at raising achievement than non-profit management?

The distinction between for-profit and non-profit management has been a topic of continuing discussion in the scholarly literature on school reform. Friedman (1955) and Coulson (1999, pp. 304-306) have theorized that for-profit firms will be more effective because they have clear economic incentives to lift student performance. The firm can build its reputation (and in the long run generate a profit) only if it becomes known for running effective schools. Others have suggested, however, that for-profit firms are likely to cut costs and short-change students in order to benefit the firm's owners and shareholders (Hochschild and Scovronick 2003, pp. 120-21; Levin 2001). The debate over non-profit organizations takes a different form. Some have argued that non-profit managers are likely to be effective because they have close ties to the community within which they are embedded and can enlist the energies of committed

entrepreneurs, who devote all available resources to enhance student performance (Brandl 1998; 2006). But others caution that non-profit managers may have neither the experience, resources, nor economic incentives necessary for building quality educational institutions (Hassel 2003, pp. 190-93; Merrifield 2001, pp. 32-35).

Most of the discussion of the impact of for-profit and non-profit management is theoretical, as very few systematic evaluations of the impact of either type of provider have been undertaken. In this paper, we add to the discussion by providing empirical information about the impact of for-profit and non-profit managers on student achievement in Philadelphia.

The interventions in Philadelphia do not provide an ideal setting for ascertaining the effectiveness of the two types of private management. The Philadelphia school district placed many restrictions on both for-profit and non-profit managers, and neither group was assigned a representative sample of schools to manage. Also, despite the substantial number of student observations, those in the treated condition are clustered within a fairly small number of schools assigned to the for-profit and non-profit managers, making it impossible to detect any other than fairly large management impacts on student achievement at conventional levels of statistical significance. To be confident of a result, it is a research convention that standard errors relative to the size of the impact must be small enough such that one would expect positive results at least 90, and preferably 95, times out of every 100 similar interventions. When students are observed within just a few schools, only effects that are substantively quite large will be significantly different from a null effect. Despite those data limitations, useful information can still be gleaned from student test-score performance over the first four years following the introduction of the management reforms.

To estimate the impact of the intervention on student achievement, we tracked the performance of individual students in math and reading from 2001 to 2006. The first two years (2001 and 2002) provide us with information about student performance prior to the management intervention, while the subsequent four years (2003, 2004, 2005, and 2006) provide information about student performance after the interventions had begun. The measures of test performance are taken from three tests: the Terra Nova, the Stanford 9, and the Pennsylvania State System of Assessment (PSSA), the test Pennsylvania administers to comply with the accountability requirements in the federal No Child Left Behind Act (NCLB).

Using that information, we found that students after three and four years at schools managed by for-profit managers made larger test-score gains, on average, in math than would be expected had those schools remained under district management. The impact of for-profit management on average student math performance after three and four years was quite large—approximately 26 percent and 29 percent of a standard deviation, respectively, which amounts to approximately two-thirds of an additional year of learning over this time span. However, the impact of for-profit management on average reading performance after three and four years was substantially less—only about 7 to 8 percent of a standard deviation, about one-quarter year's worth of learning, an impact that was not statistically significant.

For non-profits, we found mostly negative average impacts in both math and reading, but none of the effects were statistically significant. In the third and fourth year that negative impact was 15 percent and 16 percent of a standard deviation in math and, in year 3, it was 9 percent of a standard deviation in reading. After four years, the impact in reading was positive, at 4 percent of a standard deviation.

The impact of non-profit management on student performance in math is both substantively and statistically different from the impact of for-profit management, a difference equivalent to more than a year's worth of learning over four years, a strong sign that for-profit management has proved superior to non-profit management in Philadelphia.

## The Data

Our analysis is based on a student-level data set provided to Harvard's Program on Education Policy and Governance by the School District of Philadelphia. That data set contains test score, demographic, and school enrollment information on Philadelphia students in grades 2-8 from 2001 through 2006.[3] A random student identifier allows us to track individual students over time.

The Pennsylvania State System of Assessment (PSSA) is the primary vehicle for holding schools accountable for improving student learning in Philadelphia. But when the private management intervention began in the fall of 2002, that system of measuring school performance was still a work in progress. Only students in 5th, 8th, and 11th grades were given the PSSA test in reading and math. Not until the spring of 2005 did schools begin testing students with the PSSA in grade 3, and not until 2006 did the 3rd-grade exam become a part of the state accountability system. Grades 4, 6, and 7 were not tested until 2006. However, two other tests were given to some Philadelphia students between 2001 and 2006. In 2001 and 2002, students in grades 3, 4, and 7 were tested on the Stanford 9; between 2003 and 2005, the Terra Nova, designed by a different company, was used at most grade levels. The Stanford 9 and Terra Nova are nationally normed tests. Some students, in some years, were given more than one test. In 2006, the Terra Nova was dropped, except for students in first and second grade. The PSSA was

[3] The data set also includes school enrollment information on non-tested grades (K-12), which allowed us to determine whether a student was new to his or her school during a given year.

given to 3rd graders in 2005, but it did not become a test that was used for accountability purposes until 2006, the same year the PSSA was introduced in grades 4, 6, and 7. (Throughout this paper we refer to school years by the calendar year in which spring falls, as that is when tests are administered. Thus, 2001 refers to school year 2000-2001, and so forth.)

The administration by grade and year of the three tests described above is summarized in Table 1. In order to make maximum use of these data, we standardize scores by test, grade, and year to have a mean of zero and standard deviation of one. For students who took two tests (many 5th and 8th graders took both the PSSA and Terra Nova from 2003 to 2005) we make use of Terra Nova scores because that test was used more consistently during the time period under study.

Ideally, student performance over time should be tracked on the same test. Peterson (2007) did in fact track the test performance on the PSSA for two cohorts of 5th graders between 2005 and 2006. He found gains from private management for cohorts moving from 5th to 8th grade in both reading and mathematics. However, his analysis could only track average student performance at the school level, not changes in the performance of individual students. In this paper we are able to track individual student progress. That could be done for the requisite number of students only by using information from all available tests, as described above. Fortunately, the majority of the observations after the intervention began come from the Terra Nova, a low-stakes test not used for accountability purposes. As a result, it is unlikely that our results are affected by any "teaching to a test" used for holding schools accountable.[4]

Students are matched to their respective schools, and schools are classified as either under for-profit management, under non-profit management, or under regular district

[4] Except for the fourth year after treatment began, when most test results are from the PSSA.

management.[5] The vast majority of schools began treatment in the fall of 2002. However, two schools were contracted to Edison Schools in 2005, and another school was contracted to Victory Schools in the fall of 2003. To ensure direct comparisons of the same cohorts, we exclude those schools from our analysis. The remaining schools are classified into treatment and control groups based on their status in fall 2002, so for the small number of schools that changed status this is an "intent-to-treat" analysis.[6] The treated schools remaining in the analysis include 30 schools under for-profit management (20 by Edison Schools, five by Victory Schools, and five by Chancellor Beacon Academies) and 16 schools under non-profit management (five by Foundations, three by the University of Pennsylvania, five by Temple University, and three by Universal Companies). As one can see, the numbers of schools under for-profit and non-profit management are only 30 and 16, respectively, making it difficult to detect any statistically significant impacts unless they are large.

The average combined reading and math test scores prior to treatment at schools assigned to for-profit and non-profit managers were 0.39 and 0.28 standard deviations below the Philadelphia average, respectively, while the pre-treatment scores of the full set of 142 regular public schools were 0.18 standard deviations above the district average. Because of that disparity, we limited the schools included in the control group to the lower half of all regular

[5] At the time schools were given to private managers, other schools in Philadelphia, known as restructured schools, were reorganized by the school district itself, and still other schools, known as the "sweet sixteen," were given additional resources. We include these schools in our analysis for added efficiency, but those schools are not part of the control group.

[6] The Chancellor Beacon contract with the district was cancelled after one year and these five schools were returned to district management. Following the method used in Gill et al. (2007), in our analysis, they remain in the "treated" condition. Math results remain statistically significant in years three and four and all results change by no more than a trivial amount when Chancellor Beacon schools are excluded from the analysis.

district schools. Those 71 schools had prior test scores that were 0.15 standard deviations below the district average, a level of performance much closer to those at the treated schools. Restricting the control group in this way allows us to make a cleaner, if not an exact, comparison while maintaining a sufficient number of schools to be able to detect sizeable management impacts at conventional levels of statistical significance.[7]

Table A1 shows summary statistics for students tested in both treatment years (so a gain score could be calculated) in four groups of schools: for-profit, non-profit, the low-performing regular district schools, and all district schools.[8] The half of regular district schools with the lowest pre-treatment test scores have average student characteristics that are more similar to the privately managed schools than the full set of regular district schools. However, some differences remain (for further discussion, see Appendix).

## Method

We estimate the impact of attending a for-profit or non-profit privately managed school (relative to that school had it remained under district management) using a quasi-experimental research design that employs "difference-in-differences" analysis.[9] The treatment group consists of schools managed by each type of provider, and the control group includes the regular public schools with test scores below the median for all regular district schools, as discussed above. To identify the effect of treatment, we calculate in a single estimating equation, after adjusting for

[7] See Appendix Table A.2 for results when all schools are included in the estimation.

[8] Control schools were selected on the basis of all students' test scores in the two pre-treatment years, while the summary statistics presented in Table A.1 are based on students for whom a gain score can be calculated from test scores in both pre-treatment years.

[9] See Cook and Campbell (1979, pp. 214-18). For another general reference on difference-in-differences estimation, see Meyer (1995). Two recent papers in the education literature that use this method are Dynarski (2003) and Hanushek and Wößmann (2006).

student and school fixed effects, the difference between average test-score gains at treated and control schools before and after the intervention begins. So, for example, if test-score gains at the schools treated by for-profit management were 20 percent of a standard deviation higher than gains before treatment, while comparably measured gains at the control schools were only 15 percent of a standard deviation higher, the estimated effect of for-profit management would be the difference between them, or 5 percent of a standard deviation.

Since the data allow for analysis over a longer period of time than just one year—an adequate number of students are observed up to two times before treatment and as many as four times after treatment—we are able to identify the effect of attending a privately managed school over a four-year time span.

As just stated, the model employed in this analysis estimates gains in student test-score performance at the for-profit and non-profit schools from one year to the next relative to what would have been expected had they remained under regular district management. The use of gain scores thus provides an estimate of treatment effects based on the extent to which students at each school do better or worse than would be expected, given their initial test scores. In addition, student fixed effects are included in the model, a procedure that allows for the estimation of treatment effects by comparing each student's test-score gains to his or her own performance at other points in time. This allows the model to account for changes in the composition of the schools' student populations over time that cannot be accounted for by students' observable characteristics. The model also controls for students who move from one school to another. Student movement among schools may be either structural—i.e., the student moves from an elementary to a middle school—or non-structural—e.g., the student changes school because the family moves to a different neighborhood, even while the student remains at an elementary

school. Controls are included for these two different types of school changes.[10] Specifically, treatment effects are estimated by the following model:

$$\Delta Y_{igst} = \alpha + \beta_1 \text{FP1}_{st} + \beta_2 \text{FP2}_{st} + \beta_3 \text{FP3}_{st} + \beta_4 \text{FP4}_{st} + \beta_5 \text{NP1}_{st} + \beta_6 \text{FN2}_{st} + \beta_7 \text{NP3}_{st} + \beta_8 \text{NP4}_{st} + \lambda_1 \text{SM}_{it} + \lambda_2 \text{NM}_{it} + \delta_i + \mu_s + \theta_{gt} + \varepsilon_{igst} \quad (1)$$

where $\Delta Y_{igst}$ is the decile-standardized test-score gain of student *i* in grade *g* in school *s* in year *t;* α is a constant; FP1 through FP4 are treatment dummy variables (with coefficients $\beta_1$ through $\beta_4$) indicating whether the school the student was attending in year *t* was in its first through fourth year of for-profit management (respectively); NP1 through NP4 are treatment dummy variables (with coefficients $\beta_5$ through $\beta_8$) indicating whether the school the student was attending in year *t* was in its first through fourth year of non-profit management (respectively);[11] SM and NM (with coefficients $\lambda_1$ and $\lambda_2$) are dummy variables indicating whether the student has made a structural or non-structural move, respectively, from the previous year;[12] δ is a vector of student fixed effects; μ is a vector of school fixed effects; θ is a vector of grade-by-year fixed effects; and ε is the error term.[13] The model is run separately for math and reading scores, and standard errors are adjusted for clustering by schools.[14]

[10] The performance of students who leave the Philadelphia school district cannot be tracked. Also, students who are in their first year in the Philadelphia schools cannot be included in the analysis because the prior-year scores necessary for calculating gain scores are not available.

[11] We also include sets of dummy variables identifying restructured and sweet sixteen schools in their first through fourth years of treatment.

[12] A student is determined to have made a structural move if he or she is in a different school from the previous year and his or her grade this year exceeds the maximum grade of the school attended during the previous year. Student are determined to have made a non-structural move if they are currently in a different school from the one attended the previous year and their grade level in the current year does not exceed the maximum grade level of the school attended during the previous year.

[13] The model is equivalent to one with a difference-in-differences interaction term (a dummy variable indicating whether the observation is a treated unit in the period after treatment has

We use a gain score as the dependent variable rather than controlling for the prior-year score on the right-hand side of the equation because controlling for an endogenous lagged dependent variable in a model that includes student fixed effects produces biased estimates (Angrist and Krueger 1999). One drawback of our model is that it assumes that the coefficient on the prior-year test score is equal to one when the true relationship between prior-year and current-year test scores may be otherwise. Of particular concern, students who received a low score on a test in a given year made, on average, larger gains the next year than did students who received a higher initial score. We addressed that issue by calculating decile-standardized gain scores that limit comparisons among students to just those whose initial scores fell within the same decile of the distribution (Hanushek et al. 2005). Decile-standardized gain scores are standardized to have a mean of zero and standard deviation of one within each grade, year, and decile of the prior-year test-score distribution.

We also included student fixed effects so that students are compared to themselves over time. Since estimates of treatment effects are based only on those students for whom at least two gain scores are observed, the student must have at least three test-score level observations. More than three observations are required to compute two gain scores if the three level scores are not consecutive. In our main results, there are 50,665 students for whom we can compute at least two

begun) for each treatment type (for-profit and non-profit) with this term interacted with the year dummies from the post-treatment period (except for the year dummy from the first post-treatment year). The advantage of the model used here is that it allows one to directly estimate treatment effects (and standard errors) for each year post-treatment, while the standard model with interaction terms would require the addition of main effect to the coefficient on each of the interaction terms in order to obtain an estimate of treatment effects for that year.

[14] Bertrand et al. (2004) show that this method is unlikely to produce unbiased standard errors as long as the number of units (in our case, schools) is more than ten.

gain scores in math, of whom 17,569 have two gain scores, 20,723 have three gain scores, 10,666 have four gain scores, and 1,707 have five gain scores.[15]

In addition to the data described above, we include observations from other Philadelphia students during the same time period (i.e., students for whom we only observe one gain score). By including the additional observations we increase the efficiency of the model (thereby reducing the size of its standard errors), but these students do not contribute to the size of the estimated effect of the treatment. For that reason, we report in the tables the number of observations that directly contribute to the estimate of treatment effects for each year.

## Results

Our main results are presented in Table 2. Four rows of information are provided for each estimated effect. In the first row, we report, in standard deviations, the estimated impact of attending a school under for-profit and under non-profit management on student test-score gains in reading and math in each of the four years after the intervention began. In the second row, the size of the standard error is given. As mentioned previously, standard errors are robust to the clustering of students within schools. In the third row, we report the number of students observed in the treated condition and, in the fourth, the number of students in the control group.

In reading, the impact after three and four years of for-profit management, though a non-trivial 7 to 8 percent of a standard deviation, does not differ from zero at conventional levels of statistical significance. In math, however, one can be confident that the impact of for-profit management, after three and four years, does in fact differ from zero at conventional levels of statistical significance, because the impact of for-profit management is much larger—26 percent of a standard deviation by the third year and 29 percent by the fourth. That impact grew from 12

[15] The observations for reading scores are similar in number.

percent of a standard deviation in year one to 24 percent of a standard deviation in year two, to the statistically significant impacts of 26 and 29 percent of a standard deviation in years three and four. The larger impacts in the later years may be because students have been in a positive educational situation for longer periods of time or because the for-profit private managers are gaining in experience.[16]

In contrast, seven of the eight impacts of non-profit management have a negative sign, indicating the possibility that such management had an adverse impact on student performance. The size of the negative impacts in math varied from 29 percent of a standard deviation in year one to 13 percent of a standard deviation in year two, with an impact of 16 percent of a standard deviation in year four. While the negative impacts are quite large compared to what would have been expected had the schools remained under district management, none are statistically significant (largely because the number of schools observed is quite small). In reading, impacts are a positive 4 percent in year four, though negative in the preceding three years. Only the first-year negative effect of 25 percent of a standard deviation is significantly different from a null effect.[17]

[16] Among the students who contribute to our estimates, about 80% of those who were in a privately managed school in the second through fourth post-treatment years also attended a privately managed school during the previous year. However, among those observed in the third post-treatment year (2005) only 55% were in a privately managed school two years prior (2003). Among those observed in the fourth post-treatment year (2006), 66% were in a privately managed school two years prior (2004) and 41% were in a privately managed school three years prior (2003). The average student (still among those who contribute to our estimates) in a privately managed school in 2006 had spent 2.9 years in a privately managed school. These enrollment patterns may explain why our second-year treatment effect is much larger than the one-year effect, but third- and fourth-year effects are more similar to the second-year effect.

[17] The estimated impacts of the restructured and sweet sixteen schools (not shown) are generally small and always statistically insignificant in both subjects in all years. This result differs from the large, statistically significant impacts of restructured schools reported by Gill et al. (2007),

The impact on reading performance does not differ significantly between the schools under for-profit and under non-profit management. But in math the impacts of for-profit and non-profit interventions differ by a sizeable amount. Three and four years after the intervention began, schools under for-profit management outperformed schools under non-profit management in math by 41 and 46 percent of a standard deviation, respectively. Both results are statistically significant and large, a strong sign that for-profit management proved superior to non-profit management in Philadelphia.[18]

Because the impacts of for-profit and non-profit management ran in opposite directions, the overall impact of the private management experiment was muted (Table 3). When the analysis groups together the two types of private management, impacts are positive in both math and reading, but none of the impacts are large enough to be statistically significant. In math, the overall effects of private management grew from 2 percent of a standard deviation in the first year to 15 percent in the second year to 17 percent in the third year to 19 percent in the fourth year. None of these effects are statistically significant.

The results from our quasi-experimental model are supported by additional analyses reported in the Appendix, which also show for-profit management to be effective in math and raise concerns about the effectiveness of non-profit management. The models reported there have limitations that make them less preferable ways of estimating the impact of private management. However, the fact that the general direction of our results are robust to alternative specifications enhances the confidence that may be placed in the findings reported above.

but not from the small, statistically insignificant impacts of the sweet sixteen schools estimated in that study.

[18] We test the statistical significance of differences in treatment effects between for-profit and non-profit managers using an F test of the null hypothesis of no difference between the for-profit and non-profit treatment coefficients in a given year.

## Size of the Impacts

One way of obtaining a rough sense of the size of the impacts being reported is to calculate them in terms of years of schooling. In Philadelphia, one standard deviation in test-score gains on the math and reading components of the Terra Nova (the test taken almost exclusively during three of the four years of the post-treatment period) is equivalent to 2.4 and 3.6 years of learning, respectively, from grades 2 through 8.[19] That implies that 42 percent of a standard deviation in math gains is equivalent to one year of learning within the Philadelphia school district. In reading, 28 percent of a standard deviation in gains is equivalent to one year of learning. All calculations are rough, however, as the Terra Nova was only one of three tests that were administered.

In schools under for-profit management, the positive (29 percent of a standard deviation) impact on math scores in the fourth year of the intervention is estimated to be roughly equivalent to two-thirds of a year of learning more than would have occurred had the school been under district management. It is thus a large, educationally meaningful, impact. The fourth-year reading effect (8 percent of a standard deviation), while statistically insignificant, is roughly equivalent to one-quarter of a year of learning.

[19] We use scale scores (i.e., not standardized) to make these calculations, as the Terra Nova scale scores can be compared over time (unlike standardized scores, which by definition have a mean of zero). For each pair of adjacent grades, we calculate the average difference in scale score between the years, which is the test-score gain of the average student in the Philadelphia school district who stayed on grade level (i.e., was not held back a year). We then take the standard deviation of the gain scores (recall that our results are all calculated as percentages of standard deviations of gain scores) and divide it by the mean gain score. The resulting number is the number of years of learning that is equivalent to a one-standard-deviation change in the gain score. We then take the simple average, separately for math and reading, of this calculation for each pair of grades between 2 and 8.

In schools under non-profit management, math gains in the fourth year of the intervention were observed to be more than a one-half year of learning less than expected had the schools been under district management. This effect is statistically insignificant.

Finally, the differential impact of for-profit and non-profit management is very large. After four years, students in schools under for-profit management were learning the equivalent of over a year of math more than students in schools under non-profit management, a statistically significant difference. In reading, the differences between for-profit and non-profit effects after the first year of treatment are not statistically significant.

**Differences between Reported Research and RAND Study**

A prior investigation conducted by RAND-Research for Action (RAND-RFA) reported no impact of private management on student test-score performance in either reading or mathematics (RAND-RFA, 2006). Our study differs from that study in the following respects:

1) We use information on student performance from the lower half of regular district schools, providing a tighter comparison with the treated schools. (In the Appendix, we provide results from a model that includes all regular district schools in Philadelphia in the control group.)

2) We estimate separately the impact of for-profit and non-profit managers, finding sharp contrasts between the impacts of the two types of private management.[20]

3) Our model estimates the effect of private management on gains in student performance, while the RAND-RFA model uses levels of student performance as the outcome measure.

[20] RAND-RFA does estimate impacts for specific providers, but that reduces the number of observations of treated schools to very few, making it extremely difficult to detect potential impacts.

4) Our model, though including both student and school fixed effects, does not compare students to themselves only while they are in the same school.[21] Our model thus assumes that unobserved student characteristics that are constant over time are unaffected by the school the student attends. By contrast, RAND-RFA excludes from their analysis observations of students who are observed only once in a given school. That rule requires the exclusion of a large number of students and prohibits RAND-RFA from using gain scores as their outcome measure.[22]

5) Our model controls for student movement across schools, differentiating between structural and non-structural moves. In the case of structural moves, we might particularly worry that students who move to a middle school at the start of 5th grade may fare worse than those who remain in a K-8 school and that our results will be biased if the number of middle relative to K-8 schools is greater in the treatment or control group. Our model implicitly controls for any such differences.

[21] In other words, we do not include the school * student fixed effects interaction term that is included in the RAND model.

[22] The number of students at privately managed schools that contribute to our estimates is roughly similar to the number in the RAND-RFA study in years 2, 3, and 4. However, in year 1 fewer students contribute to our estimates than in the RAND-RFA study.

## Appendix: Alternative Models

To ascertain whether our estimates of for-profit and non-profit management impacts are sensitive to the particular model we employed, two alternative models were estimated, despite our reservations about each of them. The first model provides an estimation that includes all regular district schools in Philadelphia. The second drops the decile standardization of gain scores. Results from both models confirm those reported in the main text of this paper.

**Model that Includes All District Schools (but is otherwise identical to main model)**

Following RAND-RFA, this model includes all regular district schools in Philadelphia in the control group. The model has the efficiency advantage of including a larger number of schools in the control group. However, it includes schools that have a much more advantaged student population than those attending the treated schools. There is also a greater risk that estimates will be contaminated by mean reversion, whereby schools that had lower pre-treatment scores make larger test-score gains than the control schools simply because the treated schools are "bouncing back" from a particularly unlucky year. The large difference in pre-treatment scores between treated and the full set of public schools makes this a particularly important concern.

The results from this alternative specification are presented in Table A2. As compared to results reported in Table 2, the positive impacts of for-profit management on student performance in both reading and mathematics estimated by the alternative model are substantially larger and are statistically significant in both subjects.. The larger effects estimated here (compared to our main model) suggests that our preferred results mitigate bias from mean reversion. Differences in impacts of for-profit and non-profit management are similar to those reported in Table 2, however.

**Model that Excludes Decile Standardization**

Ideally, one would like to see similar changes in test scores prior to treatment in the treated and control groups. Because the schools selected for private management were the lowest-performing schools in Philadelphia, that was not the case. To make the comparison as tight as feasible, decile-standardized scores were used in the main model. Using decile standardization improved the comparison in test-score trends between non-profit and district schools, but it did not improve the comparison between the for-profit and district schools (see the last four rows in Table A1). Given that unusual pattern, a second check on the robustness of our results was undertaken, and the results are reported in Table A3. In the model that generates those results we substitute simple standardized test-score gains for decile standardized gains while leaving the model otherwise unchanged (other than to standardize gain scores so that they have a mean of zero and a standard deviation of one so that results can be compared to those obtained from the decile-standardized model). Had the differences in trends prior to treatment significantly affected our findings, then the coefficients reported in Table A3 would differ significantly from the coefficients reported in Table 2. Inasmuch as differences are minimal, we have further reason to have confidence in our findings.

## References


Angrist, Joshua D. and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics," in the *Handbook of Labor Economics*, Vol. 3A, Orley Ashenfelter and David Card, eds., Elsevier, pp. 1278-1366.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1), pp. 249-75.

Brandl, John E. 1998. *Money and Good Intentions are Not Enough: Or Why a Liberal Democrat Thinks States Need Both Competition and Community.* Washington: Brookings Institution Press.

Brandl, John E. 2006. "Choice, Religion, Community and Educational Quality." In Robert C. Enlow and Lenore T. Ealy. *Liberty and Learning: Milton Friedman's Voucher Idea at Fifty.* Washington: CATO. pp. 25-34.

Christman, Jolley Bruce, Eva Gold, and Benjamin B. Herold. 2006. "Privatization 'Philly Style': What Can Be Learned from Philadelphia's Diverse Provider Model of School Management." 2007. *Research Brief.* Philadelphia, Pa.: Research for Action, (June). Accessed March 27, 2007: http://www.researchforaction.org/publication/details/37.

Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Coulson, Andrew J. 1999. *Market Education: The Unknown History.* New Brunswick, N. J.: Transaction.

Dynarski, Susan M. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *The American Economic Review*, 93(1), pp. 279-288.

Friedman, Milton. 1955. "The Role of Government in Education" in Robert Solow, ed. *Economics and the Public Interest.* New Brunswick, New Jersey: Rutgers University Press.

Gill, Brian, Ron Zimmer, Jolley Christman, and Suzanne Blanc. 2007. "State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia," RAND Corporation and Research for Action.

Hanushek, Eric, John Kain, Daniel O'Brien, and Steven Rivkin. 2005. "The Market for Teacher Quality." NBER Working Paper No. 11154.

Hanushek, Eric A. and Ludger Wößmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries." *The Economic Journal* 116(510), pp. C63-C76.

Hassel, Brian. 2003. "The Future of Charter Schools." In Paul E. Peterson, ed. *The Future of School Choice*. Hoover Institution Press.

Herold, Benjamin B., and Morgan Riffer. 2005. "The 'Original 86': Tracking Changes in the 'Original 86' Philadelphia Public Schools Initially Targeted for Intervention Following the 2001 State Takeover," Philadelphia, Pa.: Research for Action. Accessed October 7, 2007: http://www.researchforaction.org/publication/details/152.

Hochschild, Jennifer and Nathan Scovronick. 2003. *The American Dream and the Public Schools.* Oxford University Press.

Levin, Henry. 2001. "Bear Market: The Recent Entry of For Profit Firms into the K-12 Arena is an Intriguing Trend." *Education Next* (Spring 2001), pp. 6-15.

Mead, Sara. 2007. "Easy Way Out: Restructuring Usually Means Little has Changed," *Education Next* (Spring), pp. 52-56.

Meyer, Bruce D. 1995. "Natural and quasi-experiments in economics." *Journal of Business and Economic Statistics* 13(2), pp. 151-161.

Merrifield, John. 2001. *The School Choice Wars.* Lanham, MD: Scarecrow.

Peterson, Paul E. 2007. "School Reform in Philadelphia: A Comparison of Student Achievement at Privately-Managed Schools with Student Achievement in Other District Schools," Program on Education Policy and Governance Paper 07-03.

School District of Philadelphia. 2004. "School District of Philadelphia Outperforms State Averages for PSSA Progress." Press Release. (July 22).

**Table 1. Tests Administered by Grade and Year Tested (in Spring)**

| Grade | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| 2 | | | Terra Nova | Terra Nova | Terra Nova | Terra Nova |
| 3 | Stanford 9 | Stanford 9 | Terra Nova | Terra Nova | Terra Nova | PSSA |
| 4 | Stanford 9 | Stanford 9 | Terra Nova | Terra Nova | Terra Nova | PSSA |
| 5 | PSSA | PSSA | Terra Nova & PSSA | Terra Nova & PSSA | Terra Nova & PSSA | PSSA |
| 6 | | | Terra Nova | Terra Nova | Terra Nova | PSSA |
| 7 | Stanford 9 | Stanford 9 | Terra Nova | Terra Nova | Terra Nova | PSSA |
| 8 | PSSA | PSSA | Terra Nova & PSSA | Terra Nova & PSSA | Terra Nova & PSSA | PSSA |

Notes: The Stanford 9 and Terra Nova are nationally normed tests. The Pennsylvania State System of Assessment (PSSA) is the primary vehicle for holding schools accountable for improving student learning in Philadelphia.

**Table 2. Impact on Student Test Performance Gains in Math and Reading of Attending a School under For-Profit and Non-Profit Management**

| | | Math | | Reading | |
|---|---|---|---|---|---|
| | | For-profit | Non-profit | For-profit | Non-profit |
| One-year effect | Coefficient | 0.117 | -0.292†† | 0.014 | -0.250*† |
| | Standard Error | [0.159] | [0.198] | [0.101] | [0.133] |
| | Observations (treated) | 5,103 | 1,867 | 5,073 | 1,890 |
| | Observations (control) | 6,970 | 6,970 | 6,963 | 6,963 |
| Two-year effect | Coefficient | 0.236 | -0.127†† | 0.118 | -0.029 |
| | Standard Error | [0.149] | [0.180] | [0.095] | [0.120] |
| | Observations (treated) | 9,677 | 3,613 | 9,325 | 3,604 |
| | Observations (control) | 13,290 | 13,290 | 12,929 | 12,929 |
| Three-year effect | Coefficient | 0.264** | -0.147†† | 0.071 | -0.090 |
| | Standard Error | [0.133] | [0.182] | [0.098] | [0.143] |
| | Observations (treated) | 11,462 | 4,000 | 11,146 | 3,966 |
| | Observations (control) | 15,462 | 15,462 | 15,112 | 15,112 |
| Four-year effect | Coefficient | 0.294* | -0.161†† | 0.076 | 0.036 |
| | Standard Error | [0.161] | [0.201] | [0.104] | [0.161] |
| | Observations (treated) | 8,938 | 2,981 | 8,628 | 2,948 |
| | Observations (control) | 11,919 | 11,919 | 11,576 | 11,576 |
| Total Observations | | 193,868 | | 192,214 | |
| Number of treated schools | | 30 | 16 | 30 | 16 |
| Number of control schools | | 71 | | 71 | |
| R-squared | | 0.44 | | 0.44 | |

Notes: * significant at 10%; ** significant at 5%; †† (†) indicates that the non-profit effect is significantly different from the for-profit effect at 5% (10%); robust standard errors adjusted for clustering within schools appear in brackets. Beneath the standard errors appear "effective sample sizes," which indicate the number of students in treated and untreated schools with an available gain score in the year for which the effect is estimated who also have an available gain score for at least one additional year. Total observations includes all student*year observations in the analysis, even those that contribute only to the precision of the analysis, not the estimate of impact. The decile-standardized test-score gain is the dependent variable; controls include student fixed effects, school fixed effects, and grade*year fixed effects. For-profit and non-profit effects are estimated simultaneously but presented in adjacent columns for clarity of presentation. Untreated schools include those in the bottom half of the (school aggregate) pre-treatment test-score distribution.

**Table 3. Impact on Student Test Performance Gains in Math and Reading of Attending a Privately Managed School**

| | | Math | Reading |
|---|---|---|---|
| One-year effect | Coefficient | 0.020 | -0.049 |
| | Standard Error | [0.160] | [0.098] |
| | Observations (treated) | 6,970 | 6,963 |
| | Observations (control) | 9,179 | 9,106 |
| Two-year effect | Coefficient | 0.152 | 0.088 |
| | Standard Error | [0.148] | [0.095] |
| | Observations (treated) | 13,290 | 12,929 |
| | Observations (control) | 18,841 | 18,424 |
| Three-year effect | Coefficient | 0.168 | 0.037 |
| | Standard Error | [0.134] | [0.097] |
| | Observations (treated) | 15,462 | 15,112 |
| | Observations (control) | 21,921 | 21,640 |
| Four-year effect | Coefficient | 0.186 | 0.074 |
| | Standard Error | [0.157] | [0.102] |
| | Observations (treated) | 11,919 | 11,576 |
| | Observations (control) | 15,796 | 15,585 |
| Total Observations | | 193,868 | 192,214 |
| Number of treated schools | | 46 | 46 |
| Number of control schools | | 71 | 71 |
| R-squared | | 0.44 | 0.44 |

Notes: Robust standard errors adjusted for clustering within schools appear in brackets. Beneath the standard errors appear "effective sample sizes," which indicate the number of students in treated and untreated schools with an available gain score in the year for which the effect is estimated who also have an available gain score for at least one additional year. Total observations includes all student*year observations in the analysis, even those that contribute only to the precision of the analysis, not the estimate of impact. The decile-standardized test-score gain is the dependent variable; controls include student fixed effects, school fixed effects, and grade*year fixed effects. Untreated schools include those in the bottom half of the (school aggregate) pre-treatment test-score distribution.

**Table A1. Summary Statistics of Student Characteristics Prior to Treatment**

| | For-Profit | Non-Profit | Regular District (Lower Performing) | Regular District (All Schools) |
|---|---|---|---|---|
| White | 2.6% | 1.6% | 13.5% | 23.8% |
| Black | 75.7% | 93.3% | 66.1% | 55.7% |
| Hispanic | 19.5% | 2.4% | 14.1% | 13.6% |
| Asian | 1.9% | 2.6% | 6.0% | 6.6% |
| Other Race/Ethnicity | 0.2% | 0.1% | 0.3% | 0.3% |
| Male | 49.6% | 50.1% | 50.9% | 50.3% |
| Special Education | 13.2% | 12.6% | 14.7% | 19.6% |
| Limited English Proficient | 10.8% | 3.5% | 7.8% | 8.1% |
| Math Score (Standardized Level) | -0.40 | -0.33 | -0.10 | 0.22 |
| Reading Score (Standardized Level) | -0.37 | -0.30 | -0.09 | 0.19 |
| Math Decile-Standardized Gain | -0.22 | -0.05 | -0.03 | 0.10 |
| Reading Decile-Standardized Gain | -0.21 | -0.02 | -0.03 | 0.09 |
| Math Standardized Gain | -0.10 | 0.07 | 0.00 | 0.03 |
| Reading Standardized Gain | -0.08 | 0.09 | 0.00 | 0.02 |

Notes: Summary statistics are based on students observed in each group of schools in both years (2001 and 2002) prior to treatment (so a gain score could be calculated). Test-score levels and student characteristics are averages across both years.

**Table A2. Impact on Student Test Performance Gains in Math and Reading of Attending a School under For-Profit and Non-Profit Management (Control Group Includes All Regular District Schools)**

| | | Math | | Reading | |
|---|---|---|---|---|---|
| | | For-profit | Non-profit | For-profit | Non-profit |
| One-year effect | Coefficient | 0.255** | -0.171†† | 0.116 | -0.162†† |
| | Standard Error | [0.129] | [0.170] | [0.093] | [0.125] |
| | Observations (treated) | 5,182 | 1,905 | 5,160 | 1,926 |
| | Observations (control) | 17,755 | 17,755 | 17,655 | 17,655 |
| Two-year effect | Coefficient | 0.364*** | -0.017†† | 0.173** | 0.002† |
| | Standard Error | [0.110] | [0.151] | [0.069] | [0.104] |
| | Observations (treated) | 9,826 | 3,689 | 9,464 | 3,681 |
| | Observations (control) | 39,011 | 39,011 | 38,107 | 38,107 |
| Three-year effect | Coefficient | 0.463*** | 0.028†† | 0.158** | -0.032 |
| | Standard Error | [0.108] | [0.166] | [0.078] | [0.130] |
| | Observations (treated) | 11,584 | 4,050 | 11,264 | 4,019 |
| | Observations (control) | 44,994 | 44,994 | 44,358 | 44,358 |
| Four-year effect | Coefficient | 0.425*** | -0.064†† | 0.131 | 0.060 |
| | Standard Error | [0.132] | [0.179] | [0.085] | [0.150] |
| | Observations (treated) | 9,209 | 3,035 | 8,872 | 3,001 |
| | Observations (control) | 34,570 | 34,570 | 34,126 | 34,126 |
| Total Observations | | 286,446 | | 283,978 | |
| Number of treated schools | | 30 | 16 | 30 | 16 |
| Number of control schools | | 142 | | 142 | |
| R-squared | | 0.42 | | 0.42 | |

Notes: ** significant at 5%; *** significant at 1%; †† (†) indicates that the non-profit effect is significantly different from the for-profit effect at 5% (10%); robust standard errors adjusted for clustering within schools appear in brackets. Beneath the standard errors appear "effective sample sizes," which indicate the number of students in treated and untreated schools with an available gain score in the year for which the effect is estimated who also have an available gain score for at least one additional year. Total observations includes all student*year observations in the analysis, even those that contribute only to the precision of the analysis, not the estimate of impact. The decile-standardized test-score gain is the dependent variable; controls include student fixed effects, school fixed effects, and grade*year fixed effects. For-profit and non-profit effects are estimated simultaneously but presented in adjacent columns for clarity of presentation. Untreated schools include all regular public schools.

**Table A3. Impact on Student Test Performance Gains in Math and Reading of Attending a School under For-Profit and Non-Profit Management (Substitution of Simple Standardized Gain Scores for Decile-Standardized Gain Scores)**

| | | Math | | Reading | |
|---|---|---|---|---|---|
| | | For-profit | Non-profit | For-profit | Non-profit |
| One-year effect | Coefficient | 0.143 | -0.310†† | 0.025 | -0.234*† |
| | Standard Error | [0.164] | [0.190] | [0.111] | [0.138] |
| | Observations (treated) | 5,103 | 1,867 | 5,073 | 1,890 |
| | Observations (control) | 6,970 | 6,970 | 6,963 | 6,963 |
| Two-year effect | Coefficient | 0.271* | -0.119†† | 0.119 | -0.012 |
| | Standard Error | [0.151] | [0.172] | [0.101] | [0.118] |
| | Observations (treated) | 9,677 | 3,613 | 9,325 | 3,604 |
| | Observations (control) | 13,290 | 13,290 | 12,929 | 12,929 |
| Three-year effect | Coefficient | 0.288** | -0.145†† | 0.084 | -0.052 |
| | Standard Error | [0.135] | [0.169] | [0.102] | [0.148] |
| | Observations (treated) | 11,462 | 4,000 | 11,146 | 3,966 |
| | Observations (control) | 15,462 | 15,462 | 15,112 | 15,112 |
| Four-year effect | Coefficient | 0.300* | -0.159†† | 0.059 | 0.056 |
| | Standard Error | [0.160] | [0.188] | [0.106] | [0.163] |
| | Observations (treated) | 8,938 | 2,981 | 8,628 | 2,948 |
| | Observations (control) | 11,919 | 11,919 | 11,576 | 11,576 |
| Total Observations | | 193,868 | | 192,214 | |
| Number of treated schools | | 30 | 16 | 30 | 16 |
| Number of control schools | | 71 | | 71 | |
| R-squared | | 0.37 | | 0.37 | |

Notes: * significant at 10%; ** significant at 5%; †† (†) indicates that the non-profit effect is significantly different from the for-profit effect at 5% (10%); robust standard errors adjusted for clustering within schools appear in brackets. Beneath the standard errors appear "effective sample sizes," which indicate the number of students in treated and untreated schools with an available gain score in the year for which the effect is estimated who also have an available gain score for at least one additional year. Total observations includes all student*year observations in the analysis, even those that contribute only to the precision of the analysis, not the estimate of impact. The (standardized, but not by decile) test-score gain is the dependent variable; controls include student fixed effects, school fixed effects, and grade*year fixed effects. For-profit and non-profit effects are estimated simultaneously but presented in adjacent columns for clarity of presentation. Untreated schools include those in the bottom half of the (school aggregate) pre-treatment test-score distribution.