MPCP Longitudinal Educational Growth Study
Fourth Year Report

John F. Witte

Deven Carlson

Joshua M. Cowen

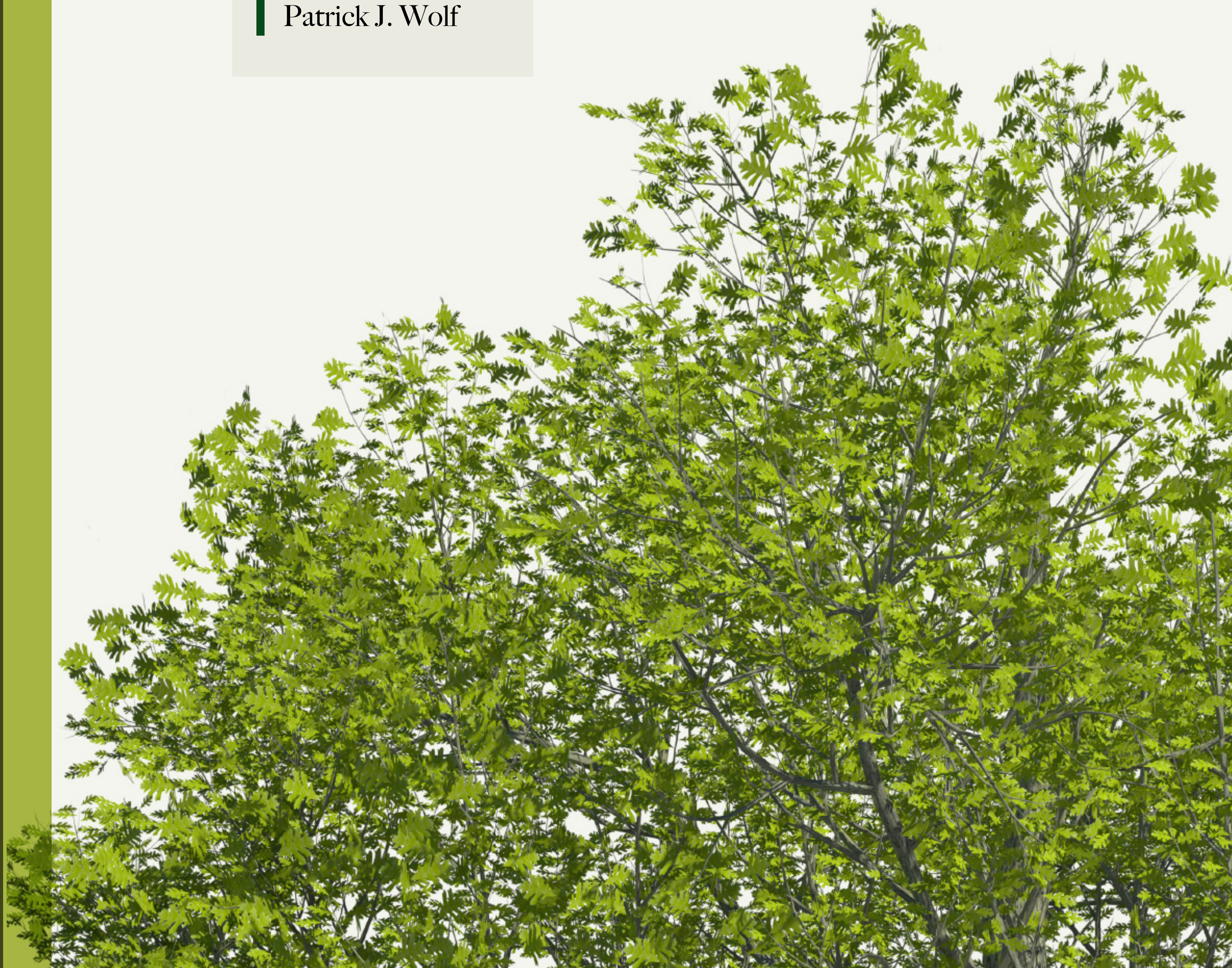
David J. Fleming

Patrick J. Wolf

SCDP Milwaukee Evaluation

Report #23

March 2011



The University of Arkansas was founded in 1871 as the flagship institution of higher education for the state of Arkansas. Established as a land grant university, its mandate was threefold: to teach students, conduct research, and perform service and outreach.

The College of Education and Health Professions established the Department of Education Reform in 2005. The department's mission is to advance education and economic development by focusing on the improvement of academic achievement in elementary and secondary schools. It conducts research and demonstration projects in five primary areas of reform: teacher quality, leadership, policy, accountability, and school choice.

The School Choice Demonstration Project (SCDP), based within the Department of Education Reform, is an education research center devoted to the non-partisan study of the effects of school choice policy and is staffed by leading school choice researchers and scholars. Led by Dr. Patrick J. Wolf, Professor of Education Reform and Endowed 21st Century Chair in School Choice, SCDP's national team of researchers, institutional research partners and staff are devoted to the rigorous evaluation of school choice programs and other school improvement efforts across the country. The SCDP is committed to raising and advancing the public's understanding of the strengths and limitations of school choice policies and programs by conducting comprehensive research on what happens to students, families, schools and communities when more parents are allowed to choose their child's school.

MPCP Longitudinal Educational Growth Study
Fourth Year Report

John F. Witte

Deven Carlson

Joshua M. Cowen

David J. Fleming

Patrick J. Wolf

SCDP Milwaukee Evaluation
Report #23
March 2011

SCHOOL CHOICE
DEMONSTRATION PROJECT

School Choice Demonstration Project
Department of Education Reform
University of Arkansas
201 Graduate Education Building
Fayetteville, AR 72701

479-575-6345

http://www.uark.edu/ua/der/SCDP/Research.html

CONTENTS:

EXECUTIVE SUMMARY

This is the fourth-year report in a five-year evaluation of the Milwaukee Parental Choice Program (MPCP). This report features analyses of student achievement growth three years after we carefully assembled longitudinal study panels of MPCP and Milwaukee Public Schools (MPS) students in 2006-07. The MPCP, which began in 1990, provides government-funded vouchers for low-income children to attend private schools in the City of Milwaukee. The maximum voucher amount in 2009-10 was $6,442, and 20,899 children used a voucher to attend either secular or religious private schools.[1] The MPCP is the oldest and largest urban school voucher program in the United States. This evaluation was authorized by 2005 Wisconsin Act 125, which was enacted in 2006.

The general purposes of the evaluation are to analyze the effectiveness of the MPCP in terms of longitudinal student achievement growth and educational attainment as measured by high school graduation and college enrollment rates. The former will be primarily accomplished by measuring and estimating student growth in achievement as measured by the Wisconsin Knowledge and Concepts Examinations (WKCE) in math and reading in grades 3 through 8 and grade 10 over a five-year period. The latter will be accomplished by following the 2006-07 8th and 9th grade cohorts over a five-year period or longer. The first report of educational attainment after four years is provided in an accompanying report (Cowen et al. 2011). The general research design for this evaluation consists of a comparison between a random sample of MPCP students and a matched sample of Milwaukee Public School students.

The February 2008 baseline report (Witte et al. 2008) presented sample means and standard deviations of student test scores in the subjects of math and reading on the November 2006 WKCE tests. The second and third year reports, released in 2009 and 2010, estimated differences in achievement growth for the MPCP and MPS samples from baseline 2006-07 achievement. The conclusions were that there were no meaningful differences between the two samples of students. In this fourth year report we present results from the November 2009 WKCE tests. These results allow us to compare three-year achievement growth for students

1 This total represents the number of students using MPCP vouchers who were enrolled on the third Friday of September in private schools that remained open for the entire 2009-2010 school year. The official third-Friday in September count of MPCP students for 2009 released by the Wisconsin Department of Public Instruction was 21,062 students.

in the MPCP, relative to three-year achievement growth for the sample of matched MPS students. We present various descriptive statistics comparing test score means and distributions for math and reading for 2006-07 (baseline year) and 2009-10 (third outcome year) for each sample. We also analyze achievement growth using several multivariate statistical techniques and models.

The primary finding in all of these comparisons is that there are no statistically significant differences in student achievement growth in either math or reading between MPCP and MPS students three years after they were carefully matched to each other. There are no statistically significant differences in either math or reading in any grade when we compare simple mean differences in achievement growth or in our multivariate models, which contain control variables for prior achievement and demographic characteristics. When we restrict the sample to only those students who have remained in either the public or private sector for all four years, we again see no statistically significant differences in math and reading achievement growth between MPCP and MPS students.

In addition to these main analyses, we also conduct several supplementary analyses to gain further insight into the relationship between student achievement and MPCP or MPS attendance. First, we conduct an analysis where we introduce a variable into our multivariate models measuring whether a student has switched schools. This analysis allows us to examine the relationship between school switching and student achievement growth, and to analyze whether any differences in student achievement growth emerge between MPS and MPCP students after controlling for school switching. Similarly, we conduct an analysis where we introduce into our multivariate model a variable measuring whether a student has ever been retained in grade. This analysis helps us understand the relationship between retention and student achievement growth, and also allows us to analyze whether any differences in student achievement growth between MPS and MPCP students emerge after controlling for student retention. These analyses demonstrate a negative relationship between student achievement growth and both school switching and student retention. However, the introduction of these variables does not change the substantive conclusion of no difference in achievement growth between MPS and MPCP students.

Finally, we analyze whether there are differences in student achievement growth between MPS and MPCP students at various points in the achievement distribution. This analysis illustrates that, *in reading*, our finding of no mean difference in achievement growth between MPS and MPCP students masks a trend where MPCP students at the lower end of the achievement distribution exhibit somewhat lower growth than their MPS counterparts while MPCP students at the higher end of the achievement distribution exhibit more growth than MPS students at similar points in the distribution.[2]

2 It is important to understand that the entire distribution of student achievement for these populations of students is skewed towards the low end of national norms, so when we say "the higher end of the achievement distribution" we are talking about the higher achieving segment of a very low-achieving population.

We also provide in Appendix B an updated assessment of missing cases, defined as students we could not locate three years after baseline. For our achievement analysis, of students we have tracked over the course of the study, 22.7 percent of the total sample drawn in 2006-07 could not be located in 2009-10, with 21 and 25 percent of MPS and MPCP panelists unable to be located, respectively. This number is considerably below our initial assumption of 20 percent sample attrition per year when we conceived sample sizes, meaning that we have a higher-powered study than expected and did not need to refresh our study sample with a new set of 3rd graders each year. In examining missing students, there were few differences in student characteristics between those missing from the MPCP or the MPS panels. As noted above, a greater number of MPCP students are missing and they were less likely to be female. There are no differences in baseline test score or race/ethnicity. To adjust for the few differences that do exist, we control for all of these variables in our multivariate models, and we use nonresponse weights that were constructed using observable baseline student characteristics in all our analyses.

Throughout the report, we describe a range of cautions and caveats; the most important being that this is the fourth year of a five-year study and that student achievement trajectories often take time to change. While presently we conclude that in general there is no significant difference between MPS students and MPCP students as measured by three years of achievement, this result may change in future analyses.

This report and its companion reports continue a series of annual reports on the Milwaukee Parental Choice Program conducted by the School Choice Demonstration Project (SCDP). An initial draft of this report was greatly improved based on comments from the SCDP Research Advisory Board and research team. All remaining errors are the responsibility of the authors alone.

This ongoing research project is being funded by a diverse set of philanthropies including the Annie E. Casey, Joyce, Kern Family, Lynde and Harry Bradley, Robertson, and Walton Family Foundations. We thank them for their generous support and acknowledge that the actual content of this report is solely the responsibility of the authors and does not necessarily reflect any official positions of the various funding organizations, the University of Wisconsin, the University of Kentucky, Furman University, the University of Arkansas, or Westat, Inc. We also express our deep gratitude to MPS, the private schools in the MPCP, and the state Department of Public Instruction for willing cooperation, advice, and assistance.

INTRODUCTION

This is the fourth report in a five-year evaluation of the Milwaukee Parental Choice Program (MPCP). This program, which began in 1990, provides government-funded vouchers for low-income children to attend private schools in the City of Milwaukee. The maximum voucher amount in 2009-10 was $6,442, and 20,899 children now use a voucher to attend either secular or religious private schools.[3] The MPCP is the oldest and largest urban school voucher program in the United States. This evaluation was authorized by the 2005 Wisconsin Act 125, which was enacted in 2006.

The general purposes of the evaluation are to analyze the effectiveness of the MPCP in terms of longitudinal student achievement growth as measured by standardized tests, and educational attainment as measured by high school graduation rates. The former will be based on estimating student achievement growth measured by the Wisconsin Knowledge and Concepts Examinations (WKCE) in math and reading in grades 3 through 8 and grade 10 over a five-year period. The latter will be accomplished by following the 2006-07 8th and 9th grade cohorts over a five-year period or longer. A report on attainment after four years is being released with this report (Cowen et al. 2011). The general research design used in this study consists of a comparison between a random sample of MPCP students and a matched sample of Milwaukee Public School (MPS) students. The procedures for obtaining that sample are briefly discussed in the next section and described in detail in Appendix B of Witte et al. (2008).

In the baseline report (Witte et al. 2008), we described baseline test scores in a number of ways. The results revealed, by design, very similar baseline scores for the MPCP and matched MPS samples on the WKCE math and reading tests. The similarity was one indicator of the success of our matching algorithm. Our second year report provided one-year growth estimates from the fall of 2006 to the fall of 2007. The essence of that report was that the achievement of students in private schools utilizing vouchers grew at the same rate in math and reading as the achievement of students in the matched-MPS sample (Witte et al. 2009). Similar results were reported for two years of achievement growth in Witte et al. (2010). In this report we present data on three-year growth in student achievement between the fall of 2006 and the fall of 2009.

Our basic analytical strategy is to first describe the *main analyses* of our longitudinal observational study. We follow that with refinements and possible explanations of the main effects with a number of *supplemental analyses*. To begin our evaluation of achievement differences between the two samples, we first provide a range of descriptive statistics on achievement growth. These include measures of central tendency, such as average gains by grade, and comparisons of the entire distribution of scores using kernel density graphs. We also use a simple but intuitively appealing method, *Somers' d* statistic, to describe the chances that MPCP students did better than MPS students in the prior three years.

3 The maximum voucher amount for the 2009-10 school year was less than the maximum voucher amount for 2008-09 ($6,607). The enrollment total represents the number of students using MPCP vouchers who were enrolled on the third Friday of September in private schools that operated throughout the 2009-2010 school year. This count differs from the Department of Public Instruction count of 21,062 students in the MPCP in 2009 because the DPI total includes some students in private schools that closed during the year.

More elaborate comparisons of main effects are made using multivariate methods in which we control for the original test score of a student in 2006-07 and a number of demographic characteristics and other independent variables. Our objective is to determine if the coefficient for the variable indicating which sector the student was in at baseline (MPCP or MPS) is significantly different from zero in the statistical sense, thereby allowing us to reject the "null hypothesis" of zero difference in gains across the two school sectors.

Because this is not a controlled experiment, some students in our panels switch from the public to the private sector or vice versa. Although we can identify these sector switchers and test them, one important research issue is the way we account for them in the long term. Should, for example, a student who begins in the MPCP sample, but after several years moves to a public school, be counted for all the years as an MPCP student? That is what is done in most medical or drug clinical trials, and that is the method we employ in our first multivariate analysis. Another way to account for that student who switched school sectors would be to simply drop the student from the analysis once the move occurs and only estimate achievement growth for those years for which the student was in their "assigned" sector, public or private. We provide a variant of that approach as an alternative analysis by estimating achievement growth for only those students who stay in the same sector for all four years. A report issued last year (Cowen et al. 2010) analyzes the characteristics of student switchers in greater detail.

In addition to the main analyses, we also perform four supplemental analyses. The purpose of the supplemental analyses is to explore what might explain the differences, or lack thereof, between the MPCP and MPS students reported as a result of our main analyses. Student mobility is a problem for all student longitudinal studies, but even more so for those conducted in high poverty areas. Mobility occurs between schools, between school districts, and through dropping out of school altogether. Mobility poses several problems and raises a number of issues. First, either dropping out of school or moving to another school district, in Wisconsin or in another state, effectively ends the acquisition of test and other data for a student. This *study attrition* reduces sample sizes and could introduce biased results if the missing cases are dissimilar on relevant variables depending on whether they are missing from the MPCP or the MPS panel. We examine this issue in Appendix B for the first four years. We correct for differences in our results by weighting the data based on the probability of remaining in the study and by including student characteristics and prior achievement as control variables in multivariate models. These strategies adequately adjust for the modest differences in attrition between the MPCP and MPS samples. Finally, in the body of the text, we analyze the effects of switching school on student achievement; the effects of retention in grade on relative achievement differences between sectors; and an analysis of differences in student achievement growth between MPS and MPCP students at various points in the achievement distribution.

The report has three basic sections. The first analyzes achievement gains from 2006 to 2009; the second offers some caveats and cautions; and the last offers a summary and a set of current conclusions. Appendix A provides descriptive statistics for variables used in our multivariate analyses. We analyze the sample attrition and describe our ongoing efforts to locate missing students in Appendix B. Appendix C provides a table comparing our original samples over time on baseline scores, taking into account attrition.

STUDENT ACHIEVEMENT GAINS: 2006 to 2009

Main Analyses

The February 2008 baseline report (Witte et al. 2008) presented sample means and standard deviations of student test scores in math and reading subjects on the November 2006 WKCE tests. We intended these statistics to provide benchmark measures of achievement current to the onset of the longitudinal study, and to serve as indicators for the success of our sample selection methodology. In this Year 4 report, we present results from the November 2009 WKCE tests as measures of student achievement growth in MPCP relative to a matched-MPS sample over a three-year growth period.

Average Math and Reading Achievement and Growth

The baseline report detailed the sample selection methodology that provides valid comparisons of MPS and MPCP students. In brief, we used students' neighborhood location, baseline test scores, and demographic information to construct the MPS sample that matched the randomly selected MPCP sample. We showed in the baseline report (Witte et al. 2008) that the MPS and MPCP samples were demonstrably similar in terms of baseline test scores and other observable characteristics. This similarity was by design. Importantly, we argued that the matching algorithm—in particular the emphasis on neighborhood location—likely accounts for unobserved characteristics that may bias comparisons of student outcomes between the two sectors. We supported this assertion in part through rich survey data collected after the matching process, which showed very similar patterns of home environment, parental education, and educational experiences for students and their parents from the same neighborhoods, regardless of whether the students were in the MPCP or the MPS (Witte et al. 2008).

Because we are confident that our matching process largely eliminated differences between the samples on factors systematically influencing student achievement, we believe that simple comparisons of Year 4 mean achievement between the sectors is a valid statistical indication of any outcome differences in student learning between the MPS and MPCP sectors by the fall of 2009. Tables 1 and 2 provide weighted mean growth in *scale scores* over three different time periods in math and reading.[4] The tables record the one-, two-, and three-year achievement growth of students who were in the original sample and had test scores in 2009 and the respective comparison year. Thus, column 1 in each table records the 2008-09 one-year growth scores; column 2 the changes from 2007-2009, indicating two-year growth; and column 3 the changes from 2006 to 2009, indicating

4 Scale scores are scores generated from basic data on the number of correct answers on a multiple choice (or other) standardized test. They fall within ranges for each grade that increase in each higher grade as tests become more complex (and the variance between students increases). They are approximately normally distributed and are integer-level measures. They are designed to measure the development of a child in each subject area and are calculated using a psychometric process called Item Response Theory or IRT.

three-year growth.[5] The sample includes students who were in grades 3-8 at baseline for whom we have WKCE achievement scores in 2009 and the respective comparison year.[6] To illustrate the interpretation of this table, consider the row of students who were in 6th grade in 2009, the first results row in the table. The first results column of the table presents the average increase in scale score by sector between 2008, when the students were in 5th grade, and 2009. Similarly, columns 2 and 3 represent the change in scale score from 4th grade in 2007 to the 6th grade in 2009 and from 3rd grade in 2006 to 6th grade in 2009.

Because of variations in grade-level ranges in scale scores that are purposely built into the test design, comparing average group-level scale scores across grades is not appropriate. For example, we cannot say that MPCP 5th graders are doing better than MPS 4th graders simply because the mean is higher for 5th graders. Fifth grade achievement is measured on a separate scale from 4th grade achievement. As a result, all comparisons must be limited to students within the same grade. The important point, however, is that the range of possible scores for each grade is the same for MPS and MPCP, so cross-sector comparisons within grades are valid.

Tables 1 and 2 display achievement growth differences between MPCP students and the matched-MPS sample. Positive numbers in the difference rows favor the MPCP students, and negative numbers favor MPS students. We break out the statistics by grade in 2009 to provide a nuanced examination of the differences. The basic conclusion is that there are hardly any statistically significant differences in these mean comparisons of achievement growth. Most differences in growth between the two sectors within grade levels converge on zero. The one exception is the 7th grade 2009 cohort which exhibits a slight statistically significant advantage in two-year reading growth for the MPCP students compared to their MPS counterparts.[7]

5 Weights were created to adjust for missing test scores. The results in this report using unweighted scores were nearly identical to those using the weighted scores. Of all the comparisons in this report only one statistic was significant in the weighted data that was not significant in the unweighted data. However, accepted research protocols call for use of weighted data in this research design.

6 A very small number of students were recorded as being in 5th grade in 2009. The results for these students who were retained in grade are not presented in Table 1 or Table 2. One of our supplementary analyses further addresses the issue of retention.

7 Given that 22 specific tests of statistical significance generated the results in Tables 1 and 2, it is entirely possible that the single finding of a significant test score difference is merely a chance discovery and not necessarily indicative of a real group difference.

Table 1. Mean Math Achievement by Grade, 2006-07 to 2009-10

Grade 2009	Group	(1) One-Year Change (08-09)		(2) Two-Year Change (07-09)		(3) Three-Year Change (06-09)	
		Mean Growth	s.e. (diff)	Mean Growth	s.e. (diff)	Mean Growth	s.e. (diff)
6	MPCP	26.6		47.9		84.5	
	MPS Matched	25.7		50.5		85.0	
	(Difference)	**0.9**	**2.9**	**-2.6**	**3.6**	**-0.5**	**3.9**
7	MPCP	30.9		51.9		71.6	
	MPS Matched	29.6		56.0		75.1	
	(Difference)	**1.3**	**2.8**	**-4.1**	**3.2**	**-3.5**	**4.0**
8	MPCP	9.4		39.1		53.7	
	MPS Matched	10.9		35.7		54.8	
	(Difference)	**-1.5**	**3.3**	**3.4**	**4.0**	**-1.1**	**4.0**
10	MPCP			19.5		14.1	
	MPS Matched			16.1		7.4	
	(Difference)			**3.4**	**5.1**	**6.7**	**4.8**

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test. Figures include only students with valid test scores in years being compared. Mean changes may not sum perfectly due to rounding. Response weights were used in calculations.

Table 2. Mean Reading Achievement by Grade, 2006-07 to 2009-10

Grade 2009	Group	(1) One-Year Change (08-09)		(2) Two-Year Change (07-09)		(3) Three-Year Change (06-09)	
		Mean Growth	s.e. (diff)	Mean Growth	s.e. (diff)	Mean Growth	s.e. (diff)
6	MPCP	14.0		20.0		27.9	
	MPS Matched	15.1		24.6		28.6	
	(Difference)	**-1.1**	**3.2**	**-4.7**	**3.4**	**-0.7**	**3.8**
7	MPCP	21.5		37.5		39.7	
	MPS Matched	18.6		30.7		37.9	
	(Difference)	**2.9**	**3.4**	**6.8***	**4.1**	**1.8**	**4.2**
8	MPCP	13.3		28.8		41.9	
	MPS Matched	10.7		29.3		42.5	
	(Difference)	2.5	3.0	-0.5	4.0	-0.6	4.1
10	MPCP			3.9		14.2	
	MPS Matched			-4.7		7.1	
	(Difference)			**8.6**	**6.1**	**7.1**	**6.0**

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test. Figures include only students with valid test scores in years being compared. Mean changes may not sum perfectly due to rounding. Response weights were used in calculations.

Somers' d

To further explore statistical differences in growth between MPCP and MPS students in a descriptive framework we use an additional method relying on ordinal data analysis. This method compares the gain score from 2006 to 2009 (by subject) for each MPCP student in a 2009 grade to the three-year gain score of each MPS student in the same grade. For each comparison, if the MPCP student had higher growth they were given a +1; if the MPS student did better, they were given a -1; if they were tied, a score of 0 was recorded. The results are then summed across all comparisons and the result is divided by the number of comparisons. The result is *Somers' d*, a nonparametric measure that represents the difference between the probability that a given MPCP student will gain more than an MPS student and the probability of the opposite occurring. We also conducted the analysis on all grades pooled for math and reading, since growth scores are on the same scale for each grade.[8] Table 3 reports the results of this analysis. Positive *Somers'd* coefficients favor MPCP students.

8 See Reynolds (1997) for a further description of this procedure.

Table 3. Somers' d Statistics for Math and Reading Growth: 2006-07 to 2009-10

Subject/Grade	Somers' D Coefficient (s.e.)
Math 5	-.22 (.31)
Math 6	-.02 (.05)
Math 7	-.04 (.05)
Math 8	-.01 (.05)
Math 10	.08 (.06)
Math All Years	-.01 (.03)
Reading 5	.02 (.33)
Reading 6	.04 (.05)
Reading 7	.06 (.05)
Reading 8	.04 (.05)
Reading 10	.06 (.06)
Reading All Years	.04 (.03)

***p<0.01, **p<0.05, *p<0.10, two-tailed. Response weights used in calculations.

For example, the coefficients in Table 3 should be interpreted as follows: the probability that an MPCP 7th grader gained more than an MPS 7th grader in reading is 0.06 (or 6%) *larger* than the probability of the reverse occurring. Most importantly, none of the probability estimates are statistically significant. That finding is consistent with the finding of no difference of means presented in Tables 1 and 2.

The Distribution of Math and Reading Growth

When describing measures of central tendency (mean differences), it is advantageous to use the basic metric of achievement tests, which in most cases is the standard scale or developmental score. These scores increase in range and mean for each succeeding grade. The reason for this is that tests cover wider areas of knowledge in higher grades and include overlapping questions from the prior grade. This means that scale scores at different grades are based on different scales: 400 may be the mean math score in grade 3, while it may be 435 in grade 4. These scores have excellent psychometric properties but do not allow direct comparisons across grades or direct understanding of effect sizes. For these reasons we construct standardized z scores from scale scores using the MPS district means and standard deviations for math and reading. For all MPS students this procedure would produce an average z-score of 0 with a standard deviation of 1.[9] Our samples may deviate from these norms at

9 We computed normalized z scores by grade level in all years for reading and math. For example, the formula for Math2007 Z score in Grade 3 would be ((Grade 3 ScaleMath2007– Grade 3 MPS district mean scale score)/(Grade 3 MPS district standard deviation)).

baseline to the extent that our study panels are comprised of students who are more educationally disadvantaged or less educationally disadvantaged than the district norm, and subsequently those data are relevant findings.

The remainder of this section analyzes the variance in student test scores in addition to the overall means. It is possible that similar mean achievement levels, or changes in those levels, could mask differences at different levels of achievement. For example, high-achieving MPCP students could outperform their matched MPS counterparts, while the opposite pattern could take place at the bottom of the achievement distribution. In computing the means, these could cancel each other out for no effect.

We examine whether this is the case graphically in Figures 1 and 2. The figures are Kernel densities, which are similar to histograms and represent estimates of the underlying probability distributions of the three-year change scores reported in the last columns of Tables 1 and 2. The figures are expressed in standardized z-scores, which were described above. As is apparent the distributions center on zero growth over the three year period. This does not mean that there were not achievement gains; it only means that these samples of students have not gained more than the larger MPS student population.

These figures provide perhaps the most concise comparisons of academic achievement growth between matched samples of MPS and MPCP students currently available. They indicate that mean growth is not only very similar between the sectors at this point in our study, but is also distributed in much the same way. The only exception is that reading growth for MPS students is slightly less variable than for MPCP students (Figure 2).[10] In general, similar frequencies of MPCP and MPS students were among the highest and lowest observed growth scores.

Figure 1. November 2006-09 Math Growth (Z-Scores) for All Students in Grades 5-8 and 10

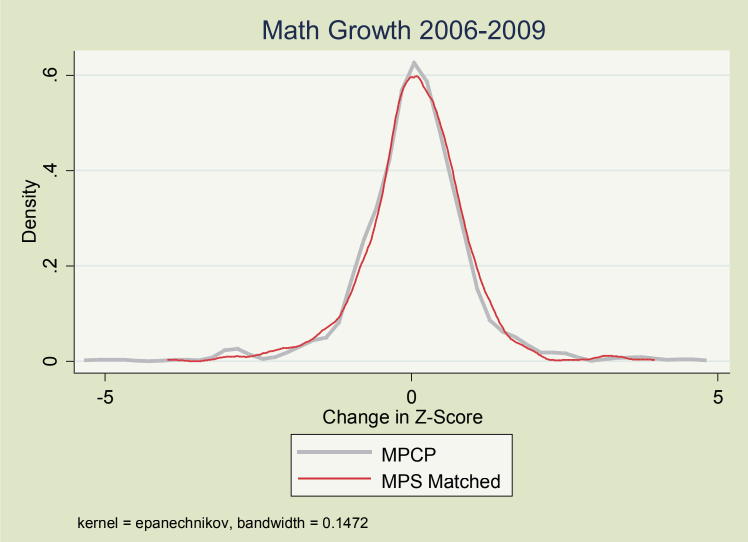
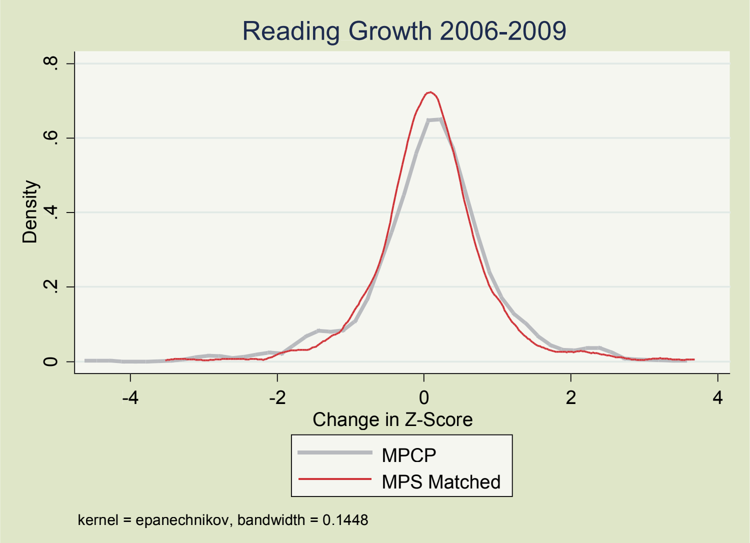


Figure 2. November 2006-09 Reading Growth (Z-Scores) for All Students in Grades 5-8 and 10



10 This is indicated by the higher spike of MPS students around the mean, which signifies less variance than for MPCP students.

Statistical Models of Math and Reading Achievement

We are confident that the strength of our matching algorithm allows us to present the above results as valid comparisons of MPCP and MPS academic achievement growth in Year 4. However, even in the context of a random assignment study—considered by many evaluators to be the "gold standard" for internal validity—there is still analytical benefit to more elaborately modeling achievement as a function of observable student baseline characteristics (e.g., Wolf et al. 2007, p. 33). In particular, the addition of a prior test score as a covariate can improve the precision of the estimate of a program effect. We formulate a simple statistical model of Year 4 achievement conditioned on baseline achievement, baseline public/private school status, and student grade level:

(eq1) $$Y_{2009,\,i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006,\,i} + \beta_3 G_i + \varepsilon_i$$

In this equation Y_{2009} is the student test score measure as a standardized z-score, β_1 represents the impact of MPCP participation (C=1), β_2 is the impact of baseline achievement, and β_3 represents a vector of grade-specific contributions to the intercept. We include grade indicator variables to capture grade-level cohort differences. With this specification, the contribution of the baseline test to the estimate of the third-year test score is unconstrained in that β_2 can take any value.[11]

Although the prior achievement variable is perhaps the most important covariate, it is not the only conceivable control variable relevant to a model of student achievement. We formulate Equation 2 as:

(eq2) $$Y_{2009,\,i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006,\,i} + \beta_3 G_i + \beta_4 X_i + \varepsilon_i$$

where β_4 represents the impact of a set of permanent student-level characteristics, X_i, specifically gender and race/ethnicity.

Results. Table 4 provides estimates of the models specified in Equations 1-2. Descriptive statistics for covariates used in Table 4 are depicted in Table A-1. The Model 1 column for math and reading reports results from an estimate of Equation 1 while the Model 2 column corresponds to estimates of Equation 2. The results in Table 4 tell a story that is very similar to the one told by the more simple comparisons presented above. Specifically, there are no statistically significant differences in either math or reading achievement growth among students in the MPCP and MPS sectors. Although the math results suggest that MPCP students might be doing somewhat worse than their MPS counterparts, this estimate is not statistically significant at a conventional level. The validity of the models is supported by the results of the estimates of the other covariates on achievement. Native American, African-American and Hispanic students score lower on average than their white counterparts—a widespread phenomenon in education research. Girls do much better than boys on reading growth, but similarly on math.

11 Some researchers have used differences in test scores as the dependent variable by subtracting the first year test score from the second. However, if we want to model achievement growth controlling for prior achievement, this has the effect of constraining the effect of prior achievement to equal 1.0, which empirically is not the true parameter. Thus, we favor the estimation model in Equation 1.

As discussed in the introduction, there are several ways to handle the fact that students switch sectors during a longitudinal study. In Table 4 we deal with sector switching by ignoring it, that is, by assuming that students remain in their initial sector for purposes of the analysis. That means that a student who switches from MPCP to MPS will "remain" in MPCP as measured by the MPCP indicator variable. Although our study is not a randomized field trial, this assumption is standard for clinical trials in medical fields. The rationale is that in the real world people will switch medicines and conditions and it is that real-world mean effect you wish to measure. In one of our supplementary analyses we introduce a variable measuring whether a student switched schools, which also includes students who switched sectors.

Table 4. Growth Models of Math and Reading Achievement, 2006-07 to 2009-10

	Model 1 - Baseline Test		Model 2 - Baseline Test, Gender & Race	
	Math 2009	Reading 2009	Math 2009	Reading 2009
MPCP06	-0.07 (0.05)	0.01 (0.04)	-0.07 (0.05)	0.01 (0.04)
2006 Score	0.58*** (0.03)	0.60*** (0.03)	0.55*** (0.03)	0.56*** (0.03)
Nat. Am.			-0.38* (0.19)	-0.33 (0.20)
Asian			0.17 (0.15)	0.16 (0.11)
Black			-0.25*** (0.07)	-0.32*** (0.07)
Hispanic			-0.10** (0.07)	-0.19*** (0.07)
Female			0.04 (0.04)	0.14*** (0.04)
Constant	-0.31 (0.30)	0.13 (0.25)	-0.08 (0.32)	0.35 (0.27)
N	1813	1815	1813	1815
R squared	0.35	0.39	0.37	0.41
F	53.67***	70.82***	41.78***	55.66***

***p<0.01, **p<0.05, *p<0.10, two-tailed. All models contain grade dummy variables; Race variables are indicator variables with "White" as the reference category. Response weights were used and students with imputed race, gender, and baseline score are included in the estimation sample. Robust standard errors clustered by school are in parentheses.

Although we accept the classical assignment logic modeled in Table 4 to a degree, we also acknowledge there is something different when you have a comparative observational study that is attempting to assess the relative achievement between schooling sectors. After all, MPCP switchers are receiving further instruction in MPS schools, and vice-versa. One way to test the sensitivity of our results to this problem is to compare only students who stay in the same sector for all years – in this case all three subsequent years. We have done that for the models estimated in Table 5. Descriptive statistics for covariates used in Table 5 are depicted in Table A-2.

Table 5. Non-Sector Switching (Stayer) Growth Models of Math and Reading Achievement, 2006-07 to 2009-10

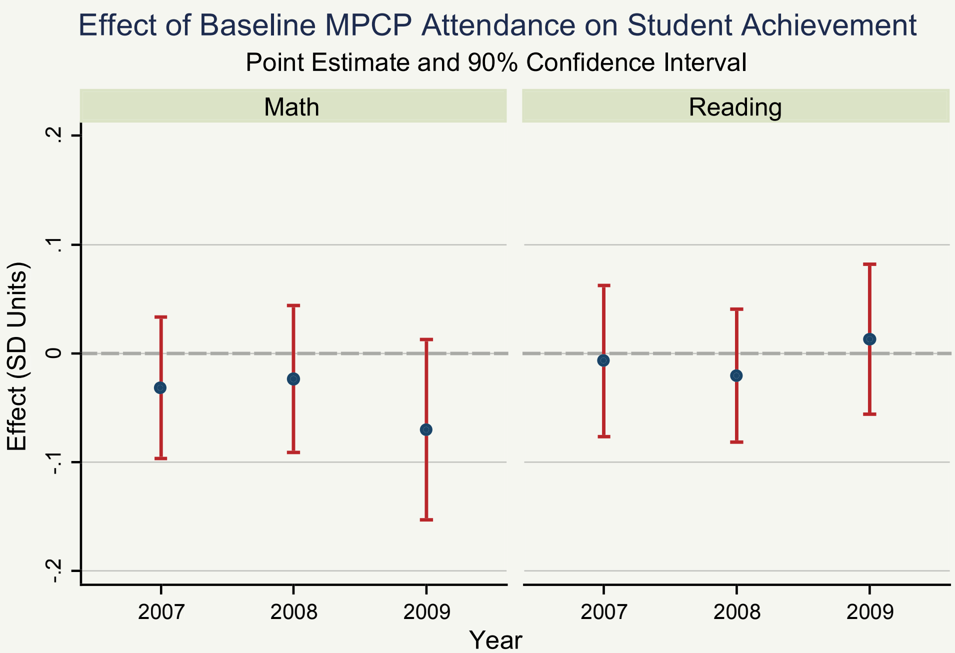
	Model 1 - Baseline Test		Model 2 - Baseline Test, Gender & Race	
	Math 2009	Reading 2009	Math 2009	Reading 2009
MPCP06	-0.06 (0.05)	0.02 (0.05)	-0.07 (0.05)	0.01 (0.05)
2006 Score	0.62*** (0.03)	0.64*** (0.03)	0.60*** (0.04)	0.61*** (0.03)
Nat. Am.			-0.44** (0.22)	-0.31 (0.23)
Asian			0.04 (0.13)	0.12 (0.10)
Black			-0.19*** (0.07)	-0.31*** (0.07)
Hispanic			-0.14* (0.08)	-0.18*** (0.07)
Female			0.03 (0.04)	0.13*** (0.05)
Constant	0.00 (0.06)	0.40 (0.19)	0.14 (0.10)	0.54*** (0.22)
N	1373	1375	1373	1375
R squared	0.39	0.42	0.40	0.44
F	76.34***	73.19***	48.40***	52.07***

***p<0.01, **p<0.05, *p<0.10, two-tailed. All models contain grade dummy variables; Race variables are indicator variables with "White" as the reference category. Response weights were used and students with imputed race, gender, and baseline score are included in the estimation sample. Robust standard errors clustered by school are in parentheses.

The results in Table 5 are nearly identical to those presented in Table 4. In math, the point estimate on the MPCP variable is negative, but this estimate does not reach a conventional level of statistical significance; for reading, the point estimates are effectively zero. Similar to Table 4, the covariates again conform to expectations.

The estimated effects for growth differences between MPS and MPCP over the three years of this study are depicted in Figure 3. This figure presents the point estimate and confidence interval for β_1 in Equation 2 (model 2, Table 4). We chose equation 2 as our preferred specification because it contains the most robust set of baseline control variables.[12] This coefficient estimates the effect of being in MPCP controlling for prior test and other baseline student characteristics. The bars indicate the statistical range the effect may take assuming a 90% (p<.1) level of statistical significance. For us to be certain at this generous level of significance that the effect is different from zero, the bars must not cross zero on the y-axis. As is apparent, MPCP students have consistently underperformed MPS students in math, but we can never be confident that the difference is statistically different from zero. For reading there is no evidence that there is any difference between the groups.

Figure 3. Mean Estimated Achievement Effect Differences for MPCP Relative to MPS-Matched Students



NOTE: Point estimates and confidence intervals based on results in Table 4, Model 2

Supplemental Analysis

In addition to the results of our main analyses, presented above, we also conduct three supplementary analyses to gain further insight into the relationship between student achievement and MPCP or MPS attendance that might explain the pattern of results uncovered in our main analyses. First, we conduct an analysis where we introduce a variable into our multivariate models measuring whether a student has switched schools. This analysis allows us to examine the relationship between school switching and student achievement growth and to analyze whether any differences in student achievement growth emerge between MPS and MPCP students after controlling for school switching. Similarly, we conduct an analysis where we introduce a variable

12 As can be seen from a comparison of Models 1 and 2 in Table 4, the selection between the models has very little impact on the graphical results.

measuring whether a student has ever been retained into our multivariate model. This analysis helps us understand the relationship between retention and student achievement growth, and also allows us to analyze whether any differences in student achievement growth between MPS and MPCP students emerge after accounting for student retention. Finally, we analyze whether there are differences in student achievement growth between MPS and MPCP students at various points in the achievement distribution.

School Switching

Previous work in different educational contexts suggests a negative impact of school switching on student outcomes (e.g. Hanushek, Kain, and Rivkin 2004; Lavertu and Witte 2008). As elaborated below, in the context of school choice in Milwaukee, MPS students switched schools *within* the MPS system at a greater rate than MPCP students switched schools within the MPCP sector. Because school switching took place *after* our matching algorithm, we could not control for it in the original design. In addition, it is possible that some school switching is in response to the quality of the school attended in either MPCP or MPS. In those cases, switching out of an MPCP or MPS school is not an independent factor in explaining student outcomes but is a consequence of enrolling in unsatisfactory MPCP or MPS schools to begin with. Thus, controlling for school switching could alter the null relationship between student achievement growth and school sector that we observed in our main analysis. We formulate equation 3 as

(eq3) $$Y_{2009,\,i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006,\,i} + \beta_3 G_i + \beta_4 X_i + \beta_5 Sch_i + \varepsilon_i$$

where β_5 represents the impact of switching schools ($Sch_i = 1$).

The results from the estimation of equation 3 are presented in columns one and two of Table 6. They confirm that school switching is negatively related to student achievement growth. Students who are not confirmed as being in their initial school exhibit lower growth in both reading and math; the precise coefficients are -0.11 and -0.09 and the results are statistically significant. However, introducing the switching variable into the model does not alter the relationship between MPCP attendance and student achievement growth; there are no statistically significant differences in student achievement growth between MPCP and MPS students when equation 3 is estimated.

Student Retention in Grade

Our data indicate that there are substantial differences in student retention between MPCP and MPS. Specifically, the data suggest that over 20 percent of students enrolled in MPS at baseline had been retained in grade at least once between the 2006-07 school year and the 2009-10 school year while only about 10 percent of students enrolled in MPCP at baseline had been retained at least once over this time period. Given this differential rate of retention, it is possible that introducing a variable measuring retention into our multivariate model could induce a relationship between student achievement growth and school sector that differs from the one observed in our main analysis. To investigate this possibility we formulate equation 4 as

(eq4) $$Y_{2009,\,i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006,\,i} + \beta_3 G_i + \beta_4 X_i + \beta_5 Ret_i + \varepsilon_i$$

where β_5 represents the impact of being retained ($Ret_i = 1$).

Estimation of equation 4 reveals evidence of a negative relationship between student retention and student achievement growth, but this relationship does not reach conventional levels of statistical significance. Moreover, introduction of the retention variable into the model does not alter the conclusions drawn from the main analysis regarding the relationship between student achievement growth and school sector; there is no statistically significant relationship between these two factors in either math or reading. Full results from equation 4 are presented in columns 3 and 4 of Table 6.

Table 6. Growth Models of Math and Reading Achievement with Retention and Switching, 2006-07 to 2009-10

	Switching		Retention		Switching & Retention	
	Math 2009 (1)	Reading 2009 (2)	Math 2009 (3)	Reading 2009 (4)	Math 2009 (5)	Reading 2009 (6)
MPCP06	-0.06	0.02	-0.08	0.01	-0.07	0.02
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
2006 Score	0.55***	0.55***	0.55***	0.56***	0.54***	0.55***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Nat. Am.	-0.36*	-0.31	-0.39**	-0.33	-0.37*	-0.32
	(0.20)	(0.20)	(0.19)	(0.20)	(0.20)	(0.20)
Asian	0.18	0.16	0.17	0.16	0.18	0.16
	(0.16)	(0.12)	(0.15)	(0.11)	(0.16)	(0.12)
Black	-0.23***	-0.31***	-0.25***	-0.32***	-0.23***	-0.30***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Hispanic	-0.10**	-0.19***	-0.10**	-0.19***	-0.09	-0.19***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Female	0.04	0.14***	0.04	0.14***	0.04	0.14***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Not confirmed in Initial school	-0.11***	-0.09**			-0.11**	-0.09**
	(0.04)	(0.04)			(0.04)	(0.04)
Ever retained			-0.13	-0.06	-0.13	-0.05
			(0.09)	(0.09)	(0.09)	(0.09)
Constant	0.00	0.42	-0.03	0.37	0.04	0.44
	(0.32)	(0.27)	(0.31)	(0.27)	(0.32)	(0.28)
N	1813	1815	1813	1815	1813	1815
R squared	0.37	0.41	0.37	0.41	0.37	0.41
F	40.03***	55.91***	39.11***	51.18***	37.67***	52.45***

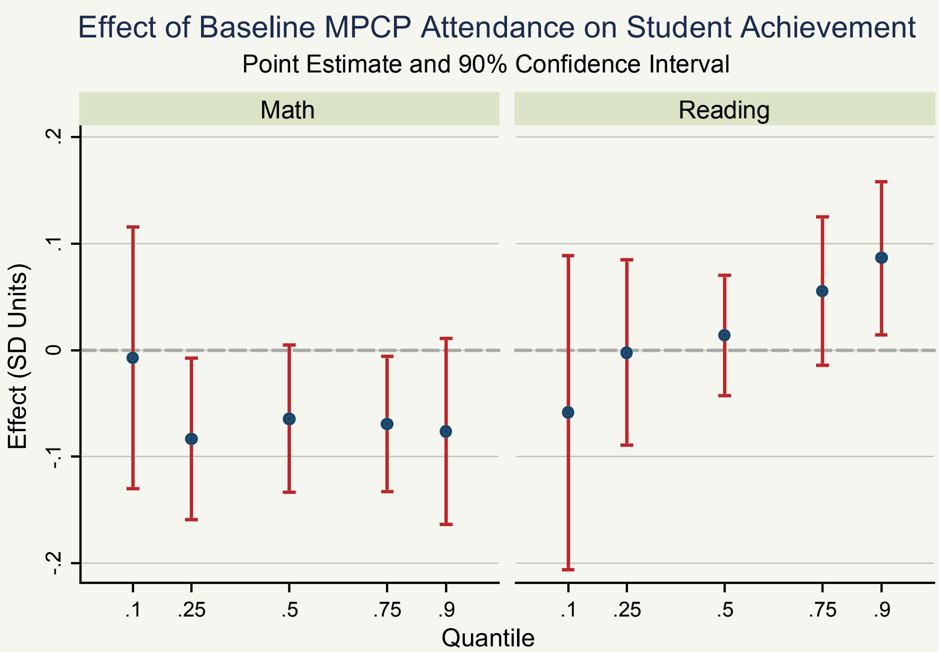
***p<0.01, **p<0.05, *p<0.10, two-tailed. All models contain grade dummy variables; Race variables are indicator variables with "White" as the reference category. Response weights were used and students with imputed race, gender, and baseline score are included in the estimation sample. Robust standard errors clustered by school are in parentheses.

We also estimated a model containing variables measuring both switching and retention, the results of which are presented in columns 5 and 6 of Table 6. The results from this model are substantively similar to those containing each variable by itself. In summary, when we include variables measuring if students either switched schools or were retained in grade, or both, there remains no significant difference in achievement growth between MPCP or MPS students.

Analysis Across the Achievement Distribution

Our main analysis illustrates that, on average, there is no difference in student achievement growth across sectors. However, these mean effects may mask interesting trends occurring across the achievement distribution. To analyze whether the relationship between MPCP attendance and student achievement growth differs by a student's position in the achievement distribution we use a technique called "quantile regression." This technique allows us to estimate the parameter or coefficient of interest (β_1 in model 2, Table 4) for students at different points in the achievement distribution. The results are depicted for both math and reading in Figure 4 below.

Figure 4. Quantile Estimated Achievement Effect Differences for MPCP Relative to MPS-Matched Student



NOTE: Point estimates and confidence intervals based on quantile regression using specification of Model 2, Table 4

Figure 4 is basically the same format and carries the same meaning as Figure 3 above. However in this case the point estimates and confidence intervals are for students at different points of the achievement distribution with respect to their 2009 test-score outcomes. Put another way, the estimates are all for achievement growth as of 2009. The results suggest that in mathematics MPCP students at all points of the achievement distribution exhibit less growth (indicated by point estimates all below zero) than their matched MPS counterparts. Note that the 25th and 75th percentiles are the only ones that produce estimates reliably different from zero, or reliably lower than those of MPS students.

The results in reading are very different and indicate that the lack of any differences in mean comparisons of MPCP and matched-MPS students masks important differences across groups of students. The trend is clear: MPCP students at the low end of the achievement outcome distribution exhibit growth that is similar to, or perhaps even slightly below, their matched MPS counterparts. However, those MPCP students at the upper levels of the achievement distribution, especially the 75th percentile or higher, achieve at higher rates of reading growth than similarly situated public school students.[13] Because the trend line of the point estimates goes from below zero to above it, the average effect is close to zero in reading as reported throughout this report (and earlier ones). It will be one objective of case studies of schools to be undertaken in 2011 to try to understand this pattern of results.

CAVEATS

These results are limited in their explanatory power in several important ways. Nearly all concern data that are missing in some way or another, either due to study attrition or because of missing or inconsistently measured information about students who remain in the study. Students who could not be located, on average, had baseline test scores that were no different from students who remained in the sample. In addition, there were no differences by gender or race. In examining missing students, there were few differences in student characteristics between those missing from the MPCP or the MPS panels. More MPCP students are missing and they are more likely to be female. There are no differences by baseline test scores or race for students missing from the MPCP and MPS samples. To adjust for the few differences that do exist, we control for these variables in our multivariate models and use nonresponse weights that were constructed using observable student characteristics in all our analyses. We also will continue to backfill missing data regarding permanent demographic characteristics of students, impute missing data on demographics that we cannot backfill, weight for missing test scores, and continue to search for missing students using a number of methods, including telephone surveys, data base searches, and even electronic tracking through the uses of "Facebook" and "Twitter."

Perhaps the most important caveat is that this study is not yet concluded with at least one more year remaining in data collection on achievement and perhaps longer for the attainment study. The achievement results in this report indicate no differences in achievement growth between public and private, voucher-receiving students. The companion report on attainment after four years (Cowen et al., 2011) comes to a somewhat different conclusion. It appears that MPCP 9th graders in 2006-07 may have graduated from high school and enrolled in four-year colleges at somewhat higher rates than did similar MPS students. It will be very important to determine if both of these results are sustained in the final year of the study.

13 It is important to understand that the entire distribution of student achievement for these populations of students is skewed towards the low end of national norms, so when we say "the higher end of the achievement distribution" we are talking about the higher achieving segment of a very low-achieving population.

SUMMARY AND CONCLUSIONS

This report presents the fourth year analysis of academic achievement in the Milwaukee Parental Choice Program (MPCP). The analysis compares a sample of MPCP students to a sample of very similar (and in most observable ways statistically identical) MPS students. A comparison of inter-sector means and other descriptive statistics did not indicate significant differences between the school sectors in terms of student achievement growth in either math or reading three years after they were carefully matched. This was also true of multivariate models that included baseline test scores, student demographic variables, and whether the student switched schools or was retained in grade. Although there was some difference in the success of MPCP schools in affecting reading achievement at the higher levels of student achievement, *the main and overwhelming conclusion is that thus far we have observed no significant difference in student achievement growth, as measured by standardized tests, between voucher-receiving private school students and a matched sample of students in MPS.*

REFERENCES

Cowen, Joshua M., David J. Fleming, John F. Witte, and Patrick J. Wolf. 2010. "School and Sector Switching in Milwaukee." SCDP Milwaukee Evaluation Report # 16, available at http://www.uark.edu/ua/der/SCDP/Milwaukee_Eval/Report_16.pdf.

Cowen, Joshua M., David J. Fleming, John F. Witte, and Patrick J. Wolf. 2011. "Student Attainment and the Milwaukee Parental Choice Program." SCDP Milwaukee Evaluation Report # 24, available at http://www.uark.edu/ua/der/SCDP/Milwaukee_Research.html

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Disruption Versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics*. 88: 1721-1746.

Howell, William G. and Paul E. Peterson, with Patrick J. Wolf and David E. Campbell. 2002. *The Education Gap: Vouchers and Urban Schools.* Washington, DC: Brookings Institution Press.

Lavertu, Stephané and Witte, John. 2008. "A Multifaceted Analysis of Milwaukee Charter Schools." Paper presented at the American Political Science Association Annual Meeting, August 28-31, Boston, MA.

Reynolds, H. T. 1977. *The Analysis of Cross-Classifications*. New York, NY. The Free Press

Witte, John F., Patrick J. Wolf, Joshua M. Cowen, David J. Fleming and Juanita Lucas-McLean. 2008. "MPCP Longitudinal Educational Growth Study Baseline Report." SCDP Milwaukee Evaluation Report #5.

Witte, John F., Patrick J. Wolf, Joshua M. Cowen, David J. Fleming and Juanita Lucas-McLean. 2009. "MPCP Longitudinal Educational Growth Study Second Year Report." SCDP Milwaukee Evaluation Report #10.

Witte, John F., Joshua M. Cowen, David J. Fleming, Patrick J. Wolf, Megan R. Condon, and Juanita Lucas-McLean. 2010. "MPCP Longitudinal Educational Growth Study Third Year Report." SCDP Milwaukee Evaluation Report #15.

Witte, John F. 2000. *The Market Approach to Education: An Analysis of America's First Voucher Program.* Princeton, NJ: Princeton University Press.

Wolf, Patrick J., Babette Guttman, Michael Puma, Lou Rizzo, Nada Eissa and Marcia Silverberg. 2007. "Evaluation of the D.C. Opportunity Scholarship Program: Impacts After One Year," Institute for Education Sciences National Center for Education Evaluation and Regional Assistance: U.S. Department of Education.

APPENDIX A

Table A-1. Descriptive Statistics for Variables Used in Table 4

	MPS Matched Counts	MPCP Counts
	N (%)	N (%)
Female	485	485
	(54.7)	(52.2)
White	93	83
	(10.5)	(8.9)
Black	569	587
	(64.2)	(63.2)
Hispanic	196*	241
	(22.2)	(25.9)
Asian	20	15
	(2.3)	(1.6)
Native	8**	2
	(0.9)	(0.2)
Switched School	277***	379
	(31.3)	(40.8)

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test. Calculations performed over the 1815 students in the estimation sample for the reading achievement models.

Table A-2. Descriptive Statistics for Variables Used in Table 5

	MPS Matched Counts	MPCP Counts
	(%)	(%)
Female	462	292
	(55.2)	(54.3)
White	90	72
	(10.8)	(13.4)
Black	529***	278
	(63.2)	(51.7)
Hispanic	190***	178
	(22.7)	(33.1)
Asian	20	9
	(2.4)	(1.7)
Native	8	1
	(1.0)	(0.2)

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test. Calculations performed over the 1159 students in the estimation sample for the reading achievement stayer model.

APPENDIX B – Study Attrition

Of the original 5,454 students in the combined MPS and MPCP panels, we were unable to locate 1,240 (23 percent) in Year 4. The rate is lower for MPS students (21 percent) compared to students who began our study in the MPCP (25 percent). Some of these students may have left Milwaukee entirely, while others may have entered independent charter schools or some other educational environment outside the scope of this report. We report these figures for the entire sample because baseline 9th graders were a part of the original match, and they are included in our study of student attainment. However, only 3,852 students (1,926 students per sector between grades 3-8) originally matched in 2006 are eligible members of our achievement study presented in this report. Of these students, we have tracked 77 percent, with only 23 percent missing after three years. That level of attrition is excellent compared to earlier studies of voucher programs (Witte 2000; Howell et al. 2002).

This appendix considers full sample attrition, or missing cases, including baseline 9th graders, who are not part of the achievement test study. There are two separate issues, differences in student characteristics of those who are missing from the study from those who are not; and differences in characteristics of missing students between sectors. Table B-1 addresses the first of these issues and Table B-2 addresses the second.

Table B-1 indicates no racial or gender differences between missing and non-missing students. Additionally, missing students have baseline test scores that are no different from the baseline scores of non-attritors. The pattern of no differences provides encouraging signs that attrition is not biasing the results of the study. However, it is possible that the characteristics of missing students varied across the MPCP and MPS sectors, a possibility that could threaten the validity of the inferences drawn in this study.

Table B-1. Sample Attrition Statistics 2006-09

	Non-Missing Students	Missing Students
Average Mean Baseline Math	-0.20	-0.20
Average Mean Baseline Reading	-0.14	-0.14
%Female	54.13	53.39
%White	8.80	8.06
%Black	66.68	67.66
%Hispanic	20.86	21.29
%Asian	3.01	2.58
%Native American	0.45	0.24
%Baseline Grade 3	13.43***	9.35
% Baseline Grade 4	12.62***	9.35
% Baseline Grade5	13.36***	9.11
% Baseline Grade 6	11.25***	15.00
% Baseline Grade 7	11.34	10.32
%Baseline Grade 8	9.21***	15.48
% Baseline Grade 9	28.79*	31.37

Stars indicate Non-missing different from missing statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test.

Table B-2 provides evidence on the difference in missing students by sector. Among students we were not able to locate at Year 4, there were no statistically significant differences in mean baseline reading or math scores between the two sectors. There were also no differences across the racial categories. However, missing MPS students were less likely to be female than missing MPCP students. There are some grade differences, as 6th and 8th graders made up a greater share of missing students for MPCP than MPS while baseline 3rd and 4th graders made up a smaller share of study attritors for MPCP than MPS; there were no differences in the proportions of 5th, 7th, and 9th graders missing between the sectors. The current study does not include a more advanced analysis of the factors associated with sample attrition (for example, a model predicting attrition that held baseline reading and grade differences constant). We do, however, weight the observations in the outcome sample by the inverse of their probability of response, given their baseline characteristics. Incorporating such sample weights into our analysis effectively recovers in our outcome sample the careful student match that we produced at baseline (e.g. Howell et al. 2002, Appendix A).

Table B-2. MPS vs. MPCP Attrition Statistics 2006-09

	MPS	MPCP
Missing Students	564 (20.68)***	676 (24.79)
Average Mean Baseline Math	-0.170	-0.222
Average Mean Baseline Reading	-0.177	-0.101
%Female	278 (49.29)***	384 (56.80)
%White	44 (7.80)	56 (8.28)
%Black	382 (67.73)	457 (67.60)
%Hispanic	117 (20.74)	147 (21.75)
%Asian	17 (3.01)	15 (2.22)
%Native American	2 (.35)	1 (.15)
%Baseline Grade 3	64 (11.35)**	52 (7.69)
% Baseline Grade 4	66 (11.70)**	50 (7.40)
% Baseline Grade5	59 (10.46)	54 (7.99)
% Baseline Grade 6	61 (10.82)***	125 (18.49)
% Baseline Grade 7	57 (10.11)	71 (10.50)
%Baseline Grade 8	76 (13.48)*	116 (17.16)
% Baseline Grade 9	181 (32.09)	208 (30.77)

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10

Table B-3 presents the status of students each post-baseline study year. The table illustrates the number of students that are confirmed as being in the same school as they were in their baseline year as well as the number of students who are in confirmed as being in their baseline sector, but in a different school. Finally, it reports the number of students who are confirmed as being in a new sector, expelled, or otherwise inaccessible.

Table B-3. Student Status, by Year: 2007-08 to 2009-10

	N (%)	Status in 2007-08		Status in 2008-09		Status in 2009-10	
		MPS	MPCP	MPS	MPCP	MPS	MPCP
1	**Same baseline sector and school**	1,526 (56.0)	1,701 (62.4)	1,459 (53.5)	978 (35.9)	1,323 (48.5)	1,099 (40.3)
2	**Same baseline sector, new school**	865 (31.7)	164 (6.0)	760 (27.9)	160 (5.9)	709 (26.0)	113 (4.1)
3	**In new sector**	71 (2.6)	402 (14.7)	124 (4.5)	457 (16.8)	105 (3.9)	801 (29.4)
4	**Expelled**	11 (0.4)	31 (1.1)	7 (0.3)	7 (0.3)	25 (0.9)	17 (0.6)
5	**Missing**	250 (9.2)	422 (15.5)	373 (13.7)	863 (31.6)	564 (20.7)	676 (24.8)
6	**Miscellaneous***	4 (0.2)	7 (0.3)	4 (0.2)	262 (9.6)	1 (0.0)	21 (0.8)
7	**Total**	2,727 (100.0)	2,727 (100.0)	2,727 (100.0)	2,727 (100.0)	2,727 (100.0)	2,727 (100.0)

* The vast majority of the 262 students in the MPCP Miscellaneous category in 2008-09 are individuals who went untracked during that year.

Appendix C. Stability of the Baseline Sample Over Time

One metric to determine how much a sample has deteriorated over time is to measure changes in the key dependent variables as attrition occurs from the sample. In our case those variables consist of 2006 math and reading scores. The issue is whether we are losing students who have nonrandom baseline scores. This measure, for example, is used by the U.S. Department of Education's *What Works Clearinghouse* to evaluate study credibility. We do not necessarily support this method, but we offer it as another way to evaluate sample attrition.

Based on the results in Table C-1, it is clear that there is very little deviation from year-to-year in the remaining students' baseline scores. The *What Works* "standard" is .25 standard deviations change from the original scores for each year and none of our estimates remotely approach that level.

Table C-1. Sector Comparisons of 2006 Baseline Scores for Students with WKCE Tests (In Z-Scores): 2007-2009

	MPCP			MPS		
Subject	N	Mean	SD	N	Mean	SD
All Students 2007:						
2006 Math Test***	1285	-0.252	0.974	1385	-0.113	0.966
2006 Reading Test	1288	-0.140	0.986	1384	-0.119	0.976
All Students 2008:						
2006 Math Test***	1126	-0.269	0.969	1257	-0.127	0.966
2006 Reading Test	1131	-0.146	0.926	1255	-0.120	0.975
All Students 2009:						
2006 Math Test***	927	-0.291	0.979	886	-0.138	0.977
2006 Reading Test	929	-0.174	0.960	886	-0.153	0.992

NOTE: *** indicates that baseline scores of MPCP students are different from baseline scores of MPS students at p < .01

About the Authors

John F. Witte is a Professor in the Department of Political Science and the Robert M. La Follette School of Public Affairs at the University of Wisconsin - Madison. He received his Ph.D. in political science from Yale University. He has been a fellow at the Russell Sage Foundation and at the Center for Advanced Study in the Behavioral Sciences. His latest book from Princeton University Press is *The Market Approach to Education: An Analysis of America's First Voucher Program*.

Deven Carlson is a doctoral candidate in the Department of Political Science and a graduate research fellow at the Wisconsin Center for Education Research. His dissertation examines how individuals' educational experiences affect democratic citizenship practices. He is a co-author of *Understanding Education Indicators: A Practical Primer for Research and Policy*, which was recently published by Teachers College Press.

Joshua M. Cowen is assistant professor of public policy at the Martin School of Public Policy and Administration at the University of Kentucky, and is a senior analyst on a new and official evaluation of the Milwaukee voucher program He has published papers on charter school sponsorship and compliance models in school voucher randomized trials, and current research includes extensions of these models in other educational settings, including science and technology interventions in different school districts. He holds a Ph.D. in political science from the University of Wisconsin-Madison, where he was among the inaugural fellows in the interdisciplinary training program in education science, funded by the U.S. Department of Education. At the Martin School, he teaches courses in public policy and program evaluation.

David J. Fleming is an Assistant Professor of Political Science at Furman University. He received his B.A. from the University of Notre Dame in 2003 and his Ph.D. from the University of Wisconsin-Madison in 2009. He was a fellow in the interdisciplinary training program in education sciences at the University of Wisconsin. His current research project examines how education policies affect parents' political and social engagement. He teaches classes in American politics, research methods, and public policy..

Patrick J. Wolf is Professor of Education Reform and 21st Century Endowed Chair in School Choice at the University of Arkansas in Fayetteville. He also is principal investigator of the School Choice Demonstration Project. Wolf has authored, co-authored, or co-edited three books and nearly 30 articles and book chapters on school choice, special education, and public management. A 1987 *summa cum laude* graduate of the University of St. Thomas (St. Paul, MN), he received his Ph.D. in Political Science from Harvard University in 1995.

SCHOOL CHOICE
DEMONSTRATION PROJECT

School Choice Demonstration Project

Department of Education Reform

University of Arkansas

201 Graduate Education Building

Fayetteville, AR 72701

479-575-6345

http://www.uark.edu/ua/der/SCDP/Milwaukee_Research.html