MPCP Longitudinal Educational Growth Study
Fifth Year Report

John F. Witte

Deven Carlson

Joshua M. Cowen

David J. Fleming

Patrick J. Wolf

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
Fifth Year Report

John F. Witte

Deven Carlson

Joshua M. Cowen

David J. Fleming

Patrick J. Wolf

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

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This is the final report in a five-year evaluation of the Milwaukee Parental Choice Program (MPCP). This report features analyses of student achievement growth four 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 2010-11 was $6,442, and 20,996 children used a voucher to attend either secular or religious private schools. 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 primary purpose of the evaluation is twofold: 1) to analyze the effectiveness of the MPCP in promoting growth in student achievement as compared to MPS; and 2) to examine the educational attainment—measured by high school graduation and college enrollment rates— of MPCP and MPS students. The first purpose is accomplished by gauging growth in student 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 for a sample of MPCP students and a carefully matched group of MPS students. The second purpose is accomplished by following the 2006-07 8th and 9th grade MPCP and matched MPS cohorts over a five-year period during which they would have had the opportunity to graduate from high school and enroll in college. An accompanying report presents the results of the attainment analysis (Cowen *et al.* 2012).

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-, third-, and fourth-year reports—released in 2009, 2010, and 2011 respectively—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 in average test-score achievement between the two samples of students. In this final, fifth-year report we present results from the November 2010 WKCE tests. These results allow us to compare four-year achievement growth for students in the MPCP sample, relative to four-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 growth for each sample. We also analyze achievement growth using several multivariate statistical techniques and models.

The primary finding that emerges from these analyses is that, for the 2010-11 school year, the students in the MPCP sample exhibit larger growth from the base year of 2006 in reading achievement than the matched MPS sample. This is the first year such an achievement growth advantage has been observed for either group in our study. Some analyses indicate that the students in the MPCP sample also exhibit larger growth in math achievement, but the results are not conclusive.

However, the results of these analyses should be interpreted with some caution. First, they represent differences that were not present in earlier years. Perhaps more important is a significant program change that took place in the final year. Beginning with the 2010-11 school year, for the first time all schools in the MPCP were required to administer the reading and math portions of the WKCE to all voucher students in grades 3-8 and 10 and publicly report the results by named schools. In effect, the MPCP schools were subjected to a test-based accountability policy for the first time. Because the test-based accountability policy was introduced after we carefully matched our sample of MPCP students to MPS students, this study is no longer solely evaluating the effectiveness of the MPCP. Rather, it is evaluating the effectiveness of both MPCP and the accountability policy that was introduced in 2010-11. There is some evidence that the larger achievement growth of the MPCP students that we observe is attributable to the introduction of the accountability policy.[1]

In addition to the main analyses, we also conduct several supplemental analyses to gain further insight into the relationship between student achievement and our matched MPCP and MPS samples. First, we analyze whether there are differences in student achievement growth between the MPS and MPCP samples at various points in the achievement distribution. This analysis provides some evidence that, in reading, the growth in student achievement among MPCP students at the lower end of the achievement distribution is somewhat larger than the reading growth of their MPS counterparts. Second, we examine whether growth in achievement varies by the amount of time spent in MPCP. A significant number of students transferred from MPCP to MPS during the course of the evaluation, and we exploit this movement to examine whether the amount of time spent in MPCP is related to growth in student achievement. The results vary by the specification of the analysis, but they demonstrate that, conditional on spending any years in MPCP, spending four or five years results in greater achievement growth than spending only one or two. However, there is also evidence that, relative to spending zero years in MPCP, students who spend one, two, or three years exhibit lower achievement growth. Finally, we conduct a sensitivity analysis to assess the extent to which our analyses produce valid estimates of student achievement growth of MPCP and matched MPS students. The results of the sensitivity analysis lend confidence to our findings.

Throughout the report, we describe a range of cautions and caveats; the most important being that the introduction of the test-score accountability policy for MPCP schools means that the study environment has changed in a meaningful way. Other caveats are related to a variety of potential concerns stemming from issues of missing data. This study has exceeded expectations with respect to minimizing the amount of missing data and we have taken several steps to mitigate any bias that may be induced by the missing data that do exist. That

---

1   Further complicating this analysis is the fact that MPS was subject to a similar accountability policy from the beginning of our evaluation. The empirical issue, which is very difficult to assess, is whether that asymmetry put MPCP students at a disadvantage from the beginning or if the introduction of such a policy has more impact the first year (as in the case for MPCP) than after a number of years (as in the case of MPS).

said, it is possible that issues related to missing data are affecting the results of this analysis, and we acknowledge this possibility.

INTRODUCTION

This is the final report in a five-year evaluation of the Milwaukee Parental Choice Program (MPCP). This report features analyses of student achievement growth four 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. Although the income limit was relaxed in 2011 to allow middle-income families to participate, that program change occurred too late to affect our study. The maximum voucher amount in 2010-11 was $6,442, and 20,996 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 primary purpose of the evaluation is twofold: 1) to analyze the effectiveness of the MPCP in promoting growth in student achievement as compared to MPS; and 2) to examine the educational attainment—measured by high school graduation and college enrollment rates—of MPCP students and a matched sample of MPS students. The first purpose is accomplished by gauging growth in student 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 for a random sample of MPCP students and a carefully matched group of MPS students. The second purpose is accomplished by following the 2006-07 8th and 9th grade MPCP and matched MPS cohorts over a five-year period. An accompanying report presents the results of the attainment analysis (Cowen *et al*. 2012). The procedures for obtaining both the MPCP and matched MPS samples 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 both two and three years of achievement growth in Witte *et al*. (2010) and Witte *et al*. (2011), respectively. In this report we present data on four-year growth in student achievement between the fall of 2006 and the fall of 2010.

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

---

1   The number of students represents the official third-Friday in September count of MPCP students for 2010-11 released by the Wisconsin Department of Public Instruction.

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 student achievement grew faster than MPS student achievement in the prior four years.

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. 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 most of these sector switchers and test them, one important research issue is the way we account for them in the analyses. 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 five years. A report issued two years ago (Cowen *et al*. 2010) analyzes the characteristics of student switchers in greater detail. The findings presented in this report confirm the overall results of the Cowen *et al*. (2010) report.

In addition to the main analyses, we also perform a number of supplemental analyses. The purpose of the supplemental analyses is to explore what might explain the differences in student achievement growth between the MPCP and MPS students that we observe in our main analyses. Perhaps the most important supplemental analysis is one that is designed to examine the role that the introduction of a test-score accountability policy affecting MPCP plays in producing the observed results. As we describe in greater detail below, the results of this analysis provide some evidence that the larger achievement growth of MPCP students that we observe in this final year may be attributable to the introduction of the accountability policy. A second supplemental analysis examines whether growth in achievement varies by the amount of time spent in MPCP. We separately analyze student achievement growth for students who spend zero, one, two, three, four, and five years in MPCP. Our third supplemental analysis investigates whether there are differences in student achievement growth between MPS and MPCP students at various points in the achievement distribution. This analysis provides insight into whether MPCP attendance might be disproportionately harmful or helpful to low- versus high-achieving students. A fourth analysis uses a technical procedure to examine the extent to which our analyses produce valid estimates of student achievement growth of MPCP and matched MPS students. Specifically, that analysis provides an estimate of the ratio of selection on unobservable characteristics to selection on observable characteristics that would need to occur in order to explain the entire observed effect.

Finally, 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 the public and private sectors, 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. For our achievement analysis—as reported in Appendix B— there are 4,007 students who are members of our initial sample and who were not scheduled to graduate by 2010-11. Of these students, 32 percent (1,280 students) could not be located in 2010-11, with respectively 28 and 36 percent of MPS and MPCP panelists unable to be located.[2] 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. In examining missing students, there are few differences in student characteristics between those missing from the MPCP or the MPS panels. A greater number of MPCP students are missing and they tend to have lower baseline (2006) math scores and are less likely to be female. There are no differences in baseline reading score or race/ethnicity between students missing from the MPCP panel and students missing from the MPS panel. 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.

The report has three basic sections. The first analyzes achievement gains from 2006 to 2010; 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 attrition into account. Extensive details regarding our research methodology are available in a separate technical appendix to this report available at http://www.uark.edu/ua/der/SCDP/Milwaukee_Research.html.

STUDENT ACHIEVEMENT GAINS: 2006 to 2010

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 on the November 2006 WKCE tests. We intended these statistics to provide benchmark measures of achievement at the onset of the longitudinal study, and to serve as indicators for the success of our sample selection methodology. In this final, fifth year report, we present results from the November 2010 WKCE tests as measures of student achievement growth in MPCP relative to a matched-MPS sample over a four-year growth period.

---

2   Our attrition analysis only considers students who should have been able to be located in 2010-11 (i.e. it does not include baseline 9th graders, who should have graduated in the 2009-10 school year).

Average Math and Reading Achievement and Growth

The baseline report detailed the sample selection methodology that provides valid comparisons of students across the MPS and MPCP sectors. 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, after matching, the MPS and MPCP samples were demonstrably similar in terms of baseline test scores and other observable characteristics.[3] 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 5 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 2010. Tables 1 and 2 provide weighted mean growth in *scale scores* over four different time periods in math and reading.[4] The tables record the one-, two-, three-, and four-year achievement growth of students who were in the original sample and had test scores in 2010 and the respective comparison year. Thus, column 1 in each table records the 2009-10 one-year growth scores; column 2 the changes from 2008-2010, indicating two-year growth; column 3 the changes from 2007 to 2010, indicating three-year growth; and column 4 the changes from 2006 to 2010, indicating four-year growth.[5] The sample includes students who were in grades 3-8 at baseline for whom we have WKCE achievement scores in 2010 and the respective comparison year.[6] To illustrate the interpretation of this table, consider the row of students who were in 7th grade in 2010, the first results row in the table. The first results column of the table presents the

---

3   There were small, but statistically significant differences in math favoring MPS over MPCP in grades 3 and 4 at baseline in 2006. That may be relevant because those grades comprise two of the three grades that are the subject of this report. However, we control for those baseline scores in our multivariate models and the major effects reported below are for reading.

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. The most important characteristic of scale scores is that an increase of one scale-score point in 1st grade should represent about the same amount of additional learning as an increase of one scale-score point in any other grade.

5   Weights were created as a robustness check to adjust for missing outcome test scores. The results using unweighted scores were substantively similar to those using the weighted scores.

6   A very small number of students were recorded as being in 6th grade in 2010. 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.

average increase in scale score by sector between 2009, when the students were in 6th grade, and 2010. Similarly, columns 2, 3, and 4 represent the change in scale score from 5th grade in 2008 to the 7th grade in 2010, from 4th grade in 2007 to 7th grade in 2010, and from 3rd grade in 2006 to 7th grade in 2010.

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 8th graders are doing better than MPS 7th graders simply because the mean is higher for 8th graders. Eighth grade achievement is measured on a separate scale from 7th 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 students, so cross-sector comparisons within grades are valid.

Tables 1 and 2 display achievement growth differences between the students in the MPCP sample and the matched-MPS sample. Positive numbers in the difference rows (highlighted in bold) favor the MPCP students, and negative numbers favor the MPS students. We break out the statistics by grade in 2010 to provide a nuanced examination of the differences. The grade levels are 7th, 8th, and 10th because accountability testing takes place in grades 3-8 and 10 each year. To be in our growth analysis this year, a student had to have been in a tested grade at baseline (3-8 or 10) and also in a tested grade four years later (3-8 or 10). Barring grade retention, only students in our study who were in grades 3, 4, or 6 at baseline, and therefore were in grades 7, 8, and 10 four years later, qualified for this final year's growth analysis, a total of 1,282 students.

Examining the results in Table 1 reveals that the 2010-11 7th grade cohort of students in our MPCP sample exhibit larger growth in math achievement than their matched MPS counterparts across all four time periods we examine. The 2010-11 8th grade MPCP students are also found to have greater growth in math achievement than their MPS counterparts across all four time periods, but none of these differences are statistically significant. The 2010-11 10th grade cohort results reveal that the students in MPCP made significantly more two-year growth than their MPS counterparts; the estimated differences for three- and four-year growth are not statistically significant.

Table 2 presents the reading results, which are substantively similar to the math results. The primary difference between the two subjects is that the MPCP advantage in reading achievement growth is concentrated in the 2010-11 8th grade cohort, while the advantage in math achievement growth is primarily in the 7th grade, as described above. Taken together, Tables 1 and 2 demonstrate that students in MPCP exhibited greater growth in both reading and math achievement across most grades and time periods, although not all of the differences are statistically significant.

Table 1. Mean Math Achievement by Grade, 2006-07 to 2010-11

Grade 2010	Group	(1) One-Year Change (09-10) Mean Growth	s.e. (diff)	(2) Two-Year Change (08-10) Mean Growth	s.e. (diff)	(3) Three-Year Change (07-10) Mean Growth	s.e. (diff)	(3) Four-Year Change (06-10) Mean Growth	s.e. (diff)
7	MPCP	31.8		54.6		77.6		111.8	
	MPS Matched	18.9		43.9		68.3		104.4	
	(Difference)	**12.9*****	**2.6**	**10.7*****	**3.1**	**9.3****	**3.6**	**7.5****	**3.8**
8	MPCP	-0.4		30.7		53.0		72.5	
	MPS Matched	-5.9		22.1		45.8		63.2	
	(Difference)	**5.4**	**4.0**	**8.6**	**6.1**	**7.2**	**5.3**	**9.4**	**5.9**
10	MPCP			21.0		27.7		39.3	
	MPS Matched			9.5		31.8		35.6	
	(Difference)			**11.5****	**4.6**	**-4.1**	**4.4**	**3.7**	**6.5**

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 2010-11

Grade 2010	Group	(1) One-Year Change (09-10) Mean Growth	s.e. (diff)	(2) Two-Year Change (08-10) Mean Growth	s.e. (diff)	(3) Three-Year Change (07-10) Mean Growth	s.e. (diff)	(3) Four-Year Change (06-10) Mean Growth	s.e. (diff)
7	MPCP	21.1		38.5		41.4		52.0	
	MPS Matched	20.0		35.1		43.0		48.0	
	(Difference)	**1.1**	**3.1**	**3.4**	**3.5**	**-1.6**	**3.8**	**3.9**	**3.9**
8	MPCP	15.2		41.4		52.0		55.6	
	MPS Matched	8.1		26.7		36.9		44.3	
	(Difference)	**7.1***	**4.2**	**14.7*****	**5.1**	**15.1*****	**5.6**	**11.2****	**5.3**
10	MPCP			7.7		16.3		23.0	
	MPS Matched			-7.2		12.2		11.2	
	(Difference)			**14.9*****	**5.0**	**4.1**	**5.4**	**11.8**	**8.8**

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.

Somers' d

To further explore statistical differences in growth between MPCP and MPS students in a descriptive framework, we use an additional method of comparison relying on ordinal data analysis. This method compares the gain score from 2006 to 2010 (by subject) for each student in the MPCP sample to the four-year gain score of each student in the MPS sample. For each comparison, if the MPCP student had higher growth, then he or she is given a +1; if the MPS student did better, then the MPCP student is 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 student in MPCP will gain more than a student in the MPS sample, and the probability of the opposite occurring.[7] Table 3 reports the results of this analysis. Positive _Somers' d_ coefficients favor the MPCP sample.

Table 3. _Somers' d_ Statistics for Math and Reading Growth: 2006-07 to 2010-11

Subject/Grade	_Somers' d_ Coefficient (s.e.)
Math 7	.11** (.06)
Math 8	.11* (.07)
Math 10	.08 (.10)
Reading 7	.11* (.06)
Reading 8	.17** (.07)
Reading 10	.11 (.11)

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

The coefficients in Table 3 should be interpreted as follows: for example, the probability that a 7th grade MPCP student gained more than a 7th grade MPS student in reading is 0.11 (or 11%) _larger_ than the probability of the reverse occurring. Examining the results in Table 3 reveals that, across both subjects and all grades, the _Somers' d_ estimate is positive. This implies that, across all pairwise comparisons, an MPCP student was between 8% and 17% more likely to exhibit greater achievement gains than an MPS student. However, the results reveal that not all of these estimates are statistically significant. In both math and reading, the seventh and eighth grade cohorts' results are statistically significant, but the 10th grade cohort results are not. These results are largely consistent with the findings presented in Tables 1 and 2.

---

7   This measure does not require that test scores are interval level numbers. Some researchers are questioning the interval level assumption of standardized tests that are being altered to accommodate state standards and achieve more accurate cut points under No Child Left Behind requirements. See Ballou (2008) and Reynolds (1977) for a further description of this procedure.

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 score (a.k.a. developmental score). 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 school district means and standard deviations for math and reading. For all students attending MPS this procedure would produce an average z-score of 0 with a standard deviation of 1.0.[8] Our samples may deviate from these norms at baseline to the extent that our study panels are comprised of students who are more or less educationally advantaged than the district norm.

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 between the panels 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 overall effect.

We examine graphically whether this is the case 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 four-year change scores reported in the last columns of Tables 1 and 2. The figures are expressed in standardized z-scores as described above. As is apparent, the distributions center close to zero growth over the four-year period. This does not mean that there were not achievement gains; it only means that these samples of students have not gained much more than the average MPS student.

These figures provide perhaps the most concise comparisons of academic achievement growth between matched samples of MPS and MPCP students currently available. Figure 1 presents the distribution of math achievement growth for MPCP and MPS students. The figure reveals that the distribution of achievement growth within the MPCP sample lies a bit to the right of MPS achievement growth. This provides a graphical illustration of the findings in Tables 1 and 3 indicating that MPCP students exhibited more growth in math achievement than their counterparts in MPS. Figure 1 also reveals that there are more students in MPCP than in MPS at the upper end of the growth distribution. Figure 2 presents the distribution of reading achievement growth for the MPCP and MPS samples and the picture is quite interesting. It reveals that the distribution of reading achievement growth among MPCP students lies somewhat to the right of the distribution of growth for the MPS sample—and thus indicates higher levels of mean reading growth—but it also exhibits substantially more variability. Although there are more students in the MPCP sample exhibiting high levels of reading growth, there are also more MPCP students at the bottom of the growth distribution.

---

8   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)).

Figure 1.
November 2006-10 Math Growth (Z-Scores)
**for All Students in Grades 7, 8, and 10**
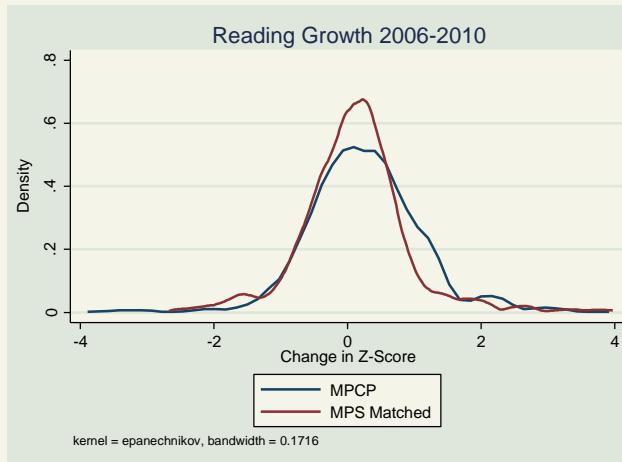
Figure 2.
November 2006-10 Reading Growth (Z-Scores)
for All Students in Grades 7, 8, and 10

Reading Growth 2006-2010

kernel = epanechnikov, bandwidth = 0.1716

Math Growth 2006-2010

kernel = epanechnikov, bandwidth = 0.1823
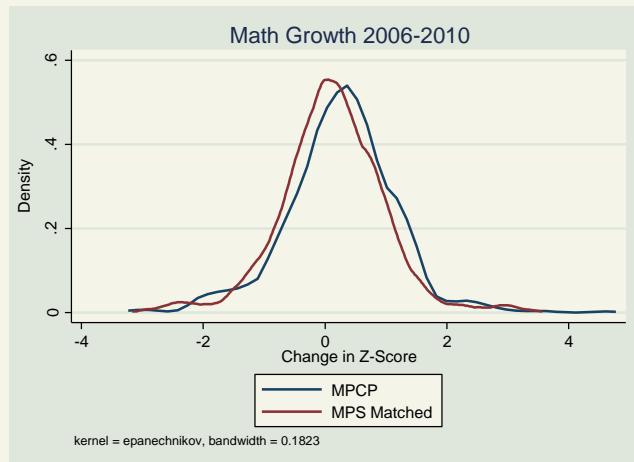
Statistical Models of Math and Reading Achievement

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 within a matched sample after four years. 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 5 achievement conditioned on baseline public/private school status, baseline achievement, and student grade level:

(eq1) $\quad Y_{2010, i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006, i} + \beta_3 G_i + \varepsilon_i$

In this equation Y_{2010} is the student test score measured 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 β_0 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 fifth-year test score is unconstrained in that β_2 can take any value.[9]

---

9   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.

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) $Y_{2010, i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006, i} + \beta_3 G_i + \beta_4 X_i + \varepsilon_i$

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

The results of equations 1 and 2 will provide insight into the average impact of baseline MPCP attendance on four-year student achievement growth, but we are also interested in knowing whether the effect of baseline MPCP attendance varies by the level of a student's baseline test score. That is, we are interested in knowing whether the effect of baseline MPCP attendance might be larger for individuals who had low baseline test scores than for individuals who had high baseline test scores, or vice versa. To address these issues, we construct a variable in which baseline MPCP attendance is interacted with (i.e. multiplied by) a student's baseline test score and include that variable in the statistical model. When the model is estimated, if this variable returns a positive and statistically significant result, then that is evidence of MPCP being more effective for students who had higher baseline test scores. If the variable returns a negative and statistically significant result, then it is evidence of the reverse occurring—students with low baseline test scores benefit more from baseline MPCP attendance than students with higher test scores at baseline. The equation for the statistical model containing the interaction of baseline MPCP attendance and baseline test scores can be written as follows:

(eq3) $Y_{2010, i} = \beta_0 + \beta_1 C_i + \beta_2 Y_{2006, i} + \beta_3 C_i {}^* Y_{2006, i} + \beta_4 G_i + \beta_5 X_i + \varepsilon_i$

Results

Table 4 provides estimates of the models specified in Equations 1-3. 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 and Model 3 columns correspond to estimates of Equation 2 and Equation 3, respectively. The results of Models 1 and 2 in Table 4 tell a story that is very similar to the one told by the more simple comparisons presented above. Specifically, the results reveal that the students in the MPCP sample exhibit greater gains in reading than their matched MPS counterparts; these estimates are statistically significant. The MPCP students are estimated to make greater gains in math as well, but the estimates are not statistically significant so we cannot rule out 0 as the true math difference in growth. The Model 3 columns present the results of the equation containing the interaction of baseline MPCP attendance with baseline test score. In reading, the negative and statistically significant coefficient on the interaction variable indicates that the positive effect of baseline MPCP attendance on four-year growth scores is higher for students with low baseline test scores than it is for students with high baseline test scores. Although the estimate on the interaction variable in math is also negative, it does not reach statistical significance. The results on the main effect of being in MPCP at baseline are not substantively or statistically affected by inclusion of the interaction term; i.e. MPCP students after four years of growth achieve higher in reading than their matched MPS counterparts.

Table 4. Growth Models of Math and Reading Achievement, 2006-07 to 2010-11

	Model 1 - Baseline Test		Model 2 - Baseline Test, Gender & Race		Model 3 - Baseline Test, Gender, Race, & Interaction	
	Math 2010	**Reading 2010**	**Math 2010**	**Reading 2010**	**Math 2010**	**Reading 2010**
MPCP06	0.08	0.15**	0.07	0.15**	0.06	0.12**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
2006 Score	0.54***	0.55***	0.50***	0.52***	0.53***	0.61***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
MPCP06 x Score06					-0.06	-0.16**
					(0.05)	(0.06)
Nat. Am.			-0.70***	-0.04	-0.69***	-0.03
			(0.17)	(0.22)	(0.17)	(0.25)
Asian			0.20**	0.26**	0.18*	0.22**
			(0.10)	(0.10)	(0.11)	(0.11)
Black			-0.34***	-0.23**	-0.35***	-0.23**
			(0.09)	(0.09)	(0.09)	(0.09)
Hispanic			-0.15	0.01	-0.16	0.00
			(0.11)	(0.09)	(0.11)	(0.09)
Female			-0.07	0.07	-0.07	0.07
			(0.07)	(0.05)	(0.07)	(0.05)
Constant	0.19	0.34**	0.47***	0.44**	0.48***	0.45**
	(0.13)	(0.17)	(0.15)	(0.20)	(0.15)	(0.20)
N	1308	1307	1308	1307	1308	1307
R squared	0.34	0.35	0.36	0.37	0.37	0.38
F	NA	NA	NA	NA	NA	NA

***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. F-tests cannot be computed using robust standard errors when a cell in the regression matrix is only a single student.

The estimated effects for growth differences between the MPS and MPCP samples over the four 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 a robust set of baseline control variables.[10] This coefficient estimates the effect of being in MPCP at baseline, controlling

---

10 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.

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, the MPCP students were estimated to grow at a similar—perhaps even slightly lower—rate in math for the first three years of the study. In the fourth year, however, the MPCP panel is found to outperform the matched MPS panel in math, although we cannot be confident that the difference between the two groups is statistically different from zero. In reading, there is again no evidence of any statistical difference in the achievement growth of the two panels for the first three years of the study. However, in the final year, the achievement growth of the students in the MPCP sample jumps, and they are estimated to outperform the MPS sample by about 0.15 standard deviations. In the following section—our supplemental analyses—we present evidence that the sudden jump in MPCP achievement from 2009 to 2010 could largely be attributable to the introduction of an accountability policy that came into effect for the 2010-11 school year.

Figure 3. Main Estimated Achievement Effect Differences for MPCP Relative to MPS-Matched Student

Point Estimate and 90% Confidence Interval
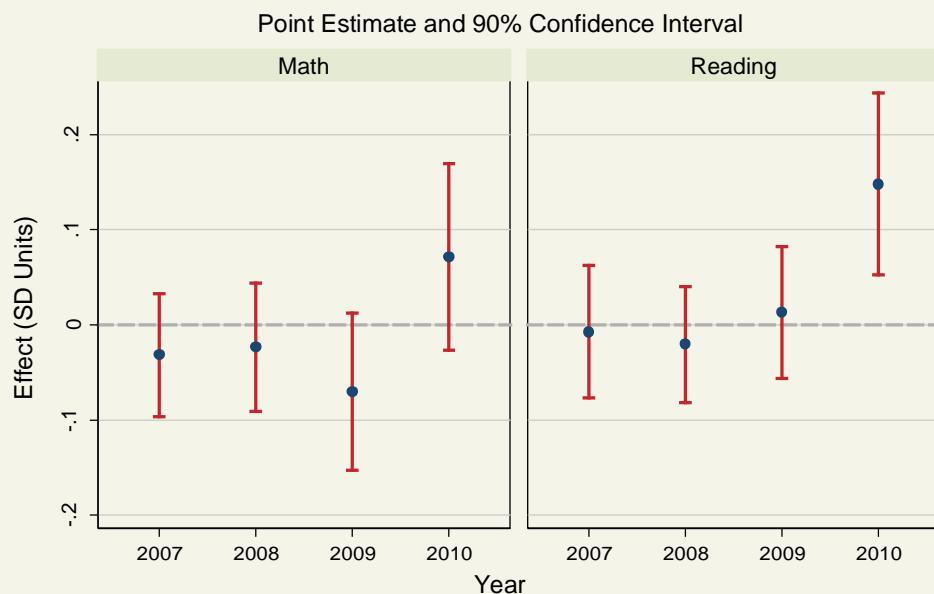
NOTE: Point estimates and confidence intervals based on results in Table 4, Model 2

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 all 1,307 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 the MPCP sample 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.

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 longitudinal observational study that is attempting to assess the relative growth in student achievement between two schooling sectors. After all, MPCP switchers are receiving further instruction in MPS schools, and vice-versa. Another way to deal with the assignment problem is only to compare those 849 students who stay in the same sector for all years—in this case all four years subsequent to baseline. Students who stayed in the MPCP or MPS across all the years of our study arguably provide us with the purest and sharpest contrast regarding the extent to which school sector influenced student achievement. We have created the models of five-year "stayers" estimated in Table 5. As indicated in a comparison of the covariate descriptive statistics for Tables 4 and 5, as depicted in Appendix Tables A-1 and A-2, the "stayers" in MPCP differ from those in our initial assignment analysis in that there are more white, Hispanic, and female students, and considerably fewer Black students who remain in MPCP after four years (Table A-2) relative to the MPCP students in our main analysis (Table A-1). An independent analysis of students who left the MPCP program to return to the public sector also showed considerable differences in baseline test scores (Cowen, *et al.* 2011). Thus, although this represents perhaps the cleanest contrast between the MPCP and MPS students in our sample, it does not necessarily provide the best estimate of the average effect of the MPCP on achievement because stayers—in both MPCP and MPS—are a distinctive subgroup of students.

Table 5. Non-Sector Switching (Stayer) Growth Models of Math and Reading Achievement, 2006-07 to 2010-11

	Model 1 - Baseline Test		Model 2 - Baseline Test, Gender & Race		Model 3 - Baseline Test, Gender, Race, & Interaction	
	Math 2010	**Reading 2010**	**Math 2010**	**Reading 2010**	**Math 2010**	**Reading 2010**
MPCP06	0.13**	0.20***	0.10*	0.17***	0.09	0.16***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
2006 Score	0.59***	0.64***	0.55***	0.61***	0.56***	0.65***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
MPCP06 x Score06					-0.04	-0.11
					(0.06)	(0.08)
Nat. Am.			-0.63***	0.16	-0.63***	0.16
			(0.19)	(0.24)	(0.18)	(0.25)
Asian			0.12	0.18*	0.11	0.16
			(0.11)	(0.10)	(0.12)	(0.10)
Black			-0.34***	-0.16*	-0.34***	-0.16**
			(0.07)	(0.08)	(0.08)	(0.08)
Hispanic			-0.19**	-0.00	-0.19**	-0.02
			(0.08)	(0.08)	(0.08)	(0.07)
Female			-0.07	0.06	-0.08	0.06
			(0.05)	(0.06)	(0.05)	(0.06)
Constant	0.20	0.51**	0.48***	0.57**	0.49***	0.59**
	(0.15)	(0.20)	(0.17)	(0.22)	(0.17)	(0.23)
N	849	851	849	851	849	851
R squared	0.45	0.43	0.47	0.44	0.47	0.44
F	92.82***	53.72***	61.80***	32.47***	57.86***	31.55***

***$p<0.01$, **$p<0.05$, *$p<0.10$, 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 substantively similar to those presented in Table 4, with a few important caveats. Models 1 and 2 illustrate that five-year MPCP attendance has a positive effect on four-year reading achievement growth. The magnitude of the estimate is in the range of 0.15-0.20 standard deviations. The results also indicate that five-year MPCP attendance has a positive effect on four-year math achievement growth. Whereas the full-sample estimates presented in Table 4 fail to reach conventional levels of statistical significance, most of the results in Table 5 do achieve that threshold. The Model 3 columns present the results of the equation containing the interaction of baseline MPCP attendance with baseline test score for four-year

stayers. Although the estimate for the interaction for stayers is negative, as it was in Table 4, it is smaller than it was for the full sample and it is no longer statistically significant.

Supplemental Analyses

In addition to the results of our main analyses, presented above, we also conduct four supplemental 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. Perhaps the most important supplemental analysis is one that is designed to examine the role that the introduction of a test-based accountability policy affecting MPCP plays in producing the observed results. As we describe in greater detail below, the results of this analysis provide some evidence that the larger achievement growth among the MPCP sample that we observe in this final year is attributable to the introduction of the accountability policy. Our second supplemental analysis examines whether growth in achievement varies by the amount of time spent in MPCP. We separately analyze student achievement growth for students who spend zero, one, two, three, four, and five years in MPCP. Our third supplemental analysis investigates whether there are differences in student achievement growth between the MPS and MPCP samples at various points in the achievement distribution. This analysis provides insight into whether MPCP attendance might be disproportionately harmful or helpful to initially low- versus high-achieving students. A fourth analysis uses a technical procedure to examine the extent to which our analyses produce valid estimates of student achievement growth of MPCP and matched MPS students. Specifically, this analysis provides an estimate of the ratio of selection on unobservable characteristics to selection on observable characteristics that would need to occur in order to explain the entire observed effect.

Introduction of Accountability Policy

Beginning with the 2010-11 school year, all schools in MPCP were required to administer the reading and math portions of the WKCE to all voucher students in grades 3-8 and 10 and provide the results to the Wisconsin Department of Public Instruction (DPI) for public reporting. The DPI was authorized to, and did, report the scores of MPCP students aggregated to the school level and named the specific schools that produced each aggregate score. In effect, the MPCP schools were subjected to a test-based accountability policy for the first time starting in the 2010-11 school year while public schools had been subjected to such an accountability policy since 2002. Because the accountability policy for MPCP was introduced after we carefully matched our sample of MPCP students to MPS students, a major new condition was introduced affecting only MPCP in the last year of the study. This makes it difficult to disentangle the achievement effects of MPCP attendance from the achievement effects of the accountability policy on MPCP students. A substantial body of previous work has shown that the introduction of test-based accountability policies has a positive impact on student achievement (Dee and Jacob 2011; Jacob 2005; Carnoy and Loeb 2002; Hanushek and Raymond 2005).

To analyze whether the greater achievement growth that we observe among the MPCP sample is more likely attributable to the effectiveness of the MPCP sector or to the introduction of the accountability policy, we perform two analyses. The first analysis is performed using all 858 students in our dataset who were confirmed as being in MPCP in 2008, 2009, and 2010. Using a statistical approach that accounts for each individual student's unobserved, time-invariant characteristics (student fixed effects) we estimate a model of achievement

that contains a series of variables indicating whether the test score was from 2008, 2009, or 2010. More formally, the model can be written as follows:

(eq4) $Y_{it} = \beta_0 + \beta_1 2009_t + \beta_2 2010_t + \gamma_i + \varepsilon_{it}$

The purpose of this approach—referred to as an "interrupted time series"—is to see whether, in comparison to previous trends, student achievement increases at a different rate after introduction of the accountability policy. In equation 4, the β_1 and β_2 coefficients should be interpreted as comparisons to student achievement in 2008. If achievement increased upon introduction of the accountability policy, then the β_2 coefficient will be substantially larger than the β_1 coefficient. Intuitively, this would show that, relative to previous achievement trends among the students in the MPCP sample, student achievement exhibited a sharp increase in the year that the accountability policy was introduced. Such a finding would certainly not provide definitive proof that the accountability policy caused the achievement increase in 2010, but it would provide suggestive evidence of that possibility.

Table 6 presents the results from the estimation of equation 4. The results illustrate that, in comparison to previous trends in the MPCP sample, student achievement increased substantially in the year that the accountability policy was introduced. In reading, student achievement is estimated to have increased by about 0.17 standard deviations in the year that the accountability policy was introduced. In math, achievement is estimated to have increased by approximately 0.18 standard deviations.

Table 6. Results of Interrupted Time Series Analysis of Implementation of Accountability Policy

	Math	Reading
Year 2009	-0.026	-0.002
	(0.023)	(0.026)
Year 2010	0.178***	0.171***
	(0.026)	(0.030)
Constant	-0.281	-0.062
	(0.013)	(0.015)
N	2002 (858 students)	2005 (858 students)
R squared	0.00	0.00
F	33.94***	21.57***

***p<0.01, **p<0.05, *p<0.10, two-tailed. Sample includes students who were confirmed in MPCP in 2008, 2009, and 2010. All models contain student fixed effects. Robust standard errors clustered by student are in parentheses. The N's in the table refer first to total observations across multiple years, and then the number of students contributing to the observations. R-squared will be low because we only include year-indicator variables in the model.

Our second supplemental analysis is slightly different from the first. Instead of comparing achievement trends among the members of the MPCP sample before and after the accountability policy was introduced, in this analysis we compare the change in achievement from 2009 to 2010 for students attending MPCP schools in those years to the change in achievement from 2009 to 2010 for students who were in MPCP schools in 2006 but had transferred to MPS schools by 2008. Due to the relatively small number of students who transferred to MPS from MPCP by 2008, we also perform this analysis using a control group consisting of all students in our dataset—not just those in the MPCP sample—who were in MPS schools in 2008, 2009, and 2010. The logic behind this approach—known as a "difference-in-differences" strategy—involves comparing the change in achievement among a group (MPCP attendees) for whom exposure to a test-based accountability system changed from 2009 (not exposed) to 2010 (exposed) to the change in achievement among a group (MPS attendees) for whom exposure to a test-based accountability policy remained constant over time (exposed in both 2009 and 2010). Although this analysis can be performed through simple subtraction, there are advantages to performing it using a statistical model.[11] Consequently, we estimate the following model:

$$\text{(eq5)} \qquad Y_{it} = \beta_0 + \beta_1 C_i + \beta_2 2010_t + \beta_3 C_i * 2010_t + \beta_4 G_i + \beta_5 X_i + \varepsilon_i$$

In this model, Y represents standardized reading or math achievement, C represents an indicator variable for MPCP attendance, and 2010 represents an indicator variable for calendar year 2010. As a result, the β_1 coefficient reflects the baseline difference in achievement between the two sectors and the β_2 coefficient reflects the time trend that is common across sectors. The interaction of those two variables—the β_3 term—reflects the differential change between MPCP and MPS and is the coefficient of interest. If this term—referred to as the "difference-in-differences" term—is positive, then it means that the change in MPCP achievement from 2009 to 2010 is greater than the change in MPS achievement over the same time period. Such a result would provide some evidence that the introduction of the accountability policy could be responsible for the jump in achievement among MPCP students observed in 2010. As was the case in equation 2, β_4 represents a vector of grade-specific contributions to the intercept and β_5 represents the impact of a set of permanent student-level characteristics, X_i, specifically gender and race/ethnicity.

Table 7 presents the results from the estimation of equation 5 for both the control group containing students who transferred from MPCP to MPS prior to 2008 and the control group containing all students attending MPS in 2008, 2009, and 2010. In both reading and math—and for both control groups—the difference-in-differences terms, identifying students being in MPCP schools in 2010, are positive and statistically significant. In reading, the coefficient has a magnitude of about 0.24 when MPCP to MPS transfers are used as the control group and about 0.09 standard deviations when all MPS students are used as the control group. In math, the

---

11   Advantages include the calculation of standard errors associated with the estimates as well as the ability to include potential time-varying confounders in the model.

respective magnitudes are each over 0.15 standard deviations.[12] Again, although this analysis does not present bulletproof evidence that the introduction of the accountability policy is entirely responsible for the increased achievement we observe in our main analyses, it does provide additional suggestive evidence that the positive MPCP achievement effects may be substantially affected by the high stakes test environment introduced in 2010.

Table 7. Results of Difference-in-Differences Analysis of Implementation of Accountability Policy

	Control group consists of students who transferred from MPCP to MPS		Control group consists of all MPS students in our data	
	Math	**Reading**	**Math**	**Reading**
MPCP	-0.080	0.135	-0.243***	0.010
	(0.086)	(0.094)	(0.040)	(0.042)
Year 2010	-0.241**	-0.090	-0.020	0.045**
	(0.098)	(0.093)	(0.021)	(0.020)
MPCP*Year 2010	0.386***	0.238**	0.158***	0.088**
	(0.110)	(0.103)	(0.040)	(0.040)
N	877 (614 students)	877 (615 students)	5543 (3,516 students)	5551 (3,521 students)
R squared	0.09	0.08	0.09	0.07
F	NA	NA	28.37***	19.04***

***p<0.01, **p<0.05, *p<0.10, two-tailed. All models contain controls for grade, race/ethnicity and gender. Robust standard errors clustered by student are in parentheses. F-tests cannot be computed using robust standard errors when a cell in the regression matrix is only a single student.

Dosage Analysis

As we describe above, our main analysis estimates the effect of being enrolled in MPCP in 2006 on subsequent growth in student achievement. We acknowledge that this approach does not explicitly account for students who switched sectors over the course of the study, so we also estimated the effect of MPCP attendance using a sample of students who remained in their baseline sector throughout the study. A third approach—the one we use in the analysis to follow—is to estimate the effect of MPCP attendance by the number of years spent in MPCP. We perform this analysis in two distinct ways. In the first way, we simply replace the variable indicating baseline MPCP attendance in equation 2 with a variable measuring the number of years spent in MPCP over

---

12   To assess whether the results we observe were driven by differential trends that were in existence prior to the introduction of the accountability policy, we performed a difference-in-differences analysis over years 2008 and 2009. Because neither sector experienced a change in exposure to an accountability policy over this time period, we expect the difference-in-differences term to not be statistically different from zero. That is exactly what we observe, and the results are available from the authors upon request.

the five years of the study; we include both a linear measure that ranges from 1-5, as well as a series of indicator variables. Our second approach takes advantage of the fact that we have information on students at each point over a five year period. This allows us to use a statistical approach that accounts for each individual student's unobserved, time-invariant characteristics when estimating the effect of MPCP attendance by the number of years spent in MPCP.

The results of these two approaches are presented in Table 8. Looking first at columns 1-2—the results of the analysis in which we replaced the variable indicating baseline MPCP attendance in equation 2 with a variable measuring the number of years spent in the MPCP—we see no statistically significant effect of additional years spent in the MPCP in math, but in reading each additional year in MPC yields a .03 standard deviation gain relative to MPS students. The analysis using year indicator variables reported in columns 3 and 4 indicate that in reading, relative to spending no years in the MPCP, spending four or five years in the program is estimated to have a positive and statistically significant effect on student achievement growth. The fifth year results in both reading and math are very similar to the "stayer" results presented in Table 5, as they should be.

Moving on to columns 5-8, which account for student's unobserved, time-invariant characteristics, we see quite different results. Specifically, these results indicate that, relative to having spent zero years in the MPCP, a student's achievement in their first, second, and third years in MPCP is actually lower, particularly in reading. The divergence in results between columns 1-4 and columns 5-8 is attributable to the fact that different analytical procedures were used. The estimates in columns 1-4 are based on all individuals who have a test score in both 2010 and 2006 while the estimates in columns 5-8 are based on students who either: 1) were in MPCP at baseline and remained there throughout the course of the study; 2) were in MPCP at baseline but switched to MPS at some point during the study; and 3) students who were in MPS at baseline but switched to MPCP at some point during the study. Students who were in MPS at baseline and remained there throughout the course of the study (stayers) are not included in the estimates presented in columns 5-8.[13] In doing this we are implicitly trading off external validity—the ability to generalize our findings to the broader population of MPCP and MPS students—for internal validity, or confidence that we are accurately estimating the effect of spending different numbers of years in MPCP instead of in MPS.

---

13   In columns 7 and 8, the zero years in MPCP reference term was identified using evidence from the years in MPS experienced by the students who started in MPS but later switched to MPCP. The results imply that these students did worse upon transferring to MPCP. The results for the one, two, and three years in MPCP terms—which are identified primarily off of students who began in MPCP and remained there—illustrate that there are no differences in achievement for students who spent one, two, three, or four years in MPCP. Readers may note that the number of observations is larger for the results presented in columns 5-8 than for columns 1-4. This is due to the fact that the unit of analysis for the results in columns 1-4 is the student while the unit of analysis for the results presented in columns 5-8 is the student-year. That is, for the results in columns 1-4, the data contain a single observation per student while the data that serves as the basis of the results in columns 5-8 may contain up to five observations per student—one each for 2006, 2007, 2008, 2009, and 2010.

Table 8. Models of Math and Reading Achievement by Number of Years Spent in MPCP, 2006-07 to 2009-10

	Four Year Growth Model- Linear		Four Year Growth Model- Indicators		Student Fixed Effects Model- Linear		Student Fixed Effects Model- Indicators	
	Math 2010 (1)	Reading 2010 (2)	Math 2010 (3)	Reading 2010 (4)	Math (5)	Reading (6)	Math (7)	Reading (8)
Yrs. in MPCP	0.02	0.03***			0.01	0.01		
	(0.01)	(0.01)			(0.01)	(0.01)		
Yr. 1 in MPCP			-0.06	-0.02			-0.09	-0.14*
			(0.09)	(0.10)			(0.07)	(0.07)
Yr. 2 in MPCP			-0.09	0.00			-0.05	-0.15**
			(0.08)	(0.09)			(0.07)	(0.08)
Yr. 3 in MPCP			0.04	0.00			-0.12	-0.14*
			(0.07)	(0.07)			(0.08)	(0.08)
Yr. 4 in MPCP			0.06	0.14*			-0.08	-0.12
			(0.08)	(0.07)			(0.08)	(0.08)
Yr. 5 in MPCP			0.08	0.16***			0.05	-0.01
β			(0.05)	(0.05)			(0.08)	(0.08)
N	1308	1307	1308	1307	13,298 (4,912 students)	13,317 (4,917 students)	13,298 (4,912 students)	13,317 (4,917 students)
R squared	0.40	0.40	0.40	0.40	0.00	0.00	0.00	0.00
F	NA	NA	NA	NA	NA	NA	NA	NA

***p<0.01, **p<0.05, *p<0.10, two-tailed. Models in columns 1-4 contain controls for baseline (2006) score, gender, race, and current (2010) grade. Columns 5-8 contain indicator variables for calendar year; Zero years in MPCP serves as the reference category for the series of indicators. Students with imputed race, gender, and baseline score are included in the estimation sample. Robust standard errors are in parentheses. F-tests cannot be computed using robust standard errors when a cell in the regression matrix is only a single student.

Analysis Across the Achievement Distribution.

Our main analysis estimates the average effect of baseline MPCP attendance on four-year achievement growth in math and reading. While informative, these mean effects may mask interesting trends occurring across the achievement distribution. To analyze whether the relationship between baseline 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 equation 2) for students at different points in the 2010 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

Point Estimate and 90% Confidence Interval
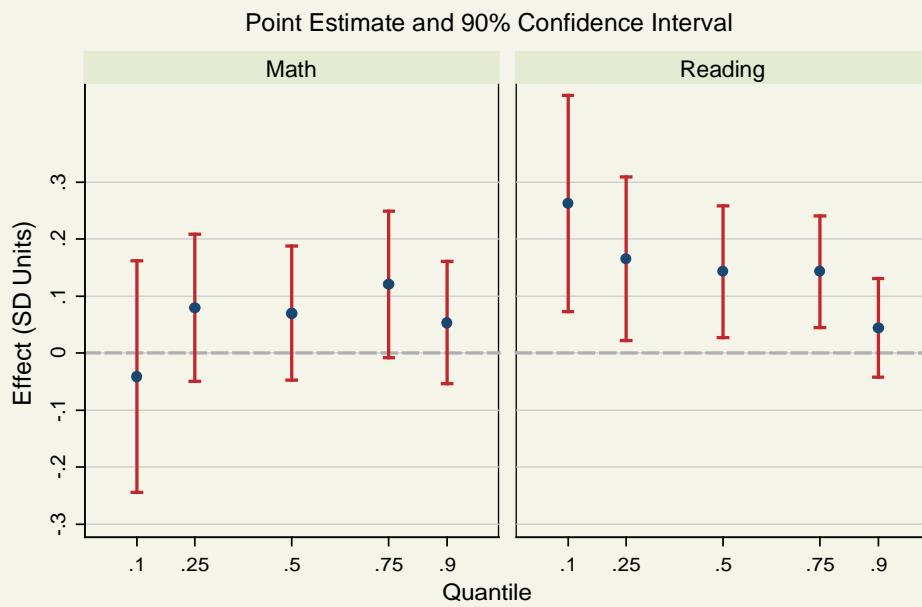
NOTE: Point estimates and confidence intervals based on quantile regression using specification of Model 2, Table 4

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 on the achievement distribution with respect to their 2010 test-score outcomes. Put another way, the estimates are all for achievement growth as of 2010. The results suggest that in math the effect of baseline MPCP attendance for students at the lower end of the achievement distribution may be somewhat smaller than the effect for students at higher points in the achievement distribution. The results in reading, however, indicate that the effect of baseline MPCP attendance on achievement growth is likely larger for students at the lower end of the distribution than for students in the 25-90th percentiles. Because all the vertical lines include the zero point, none of the results in math are significant at the 90% confidence level. On the other hand, the relationship between baseline MPCP attendance and student achievement growth in reading is positive and significant for students at every point in the distribution other than the 90th percentile.

Selection on Unobservables

A persistent concern in observational studies such as this one is the possibility that unobserved factors are driving the observed relationship between baseline MPCP attendance and student achievement growth. Altonji, Elder, and Taber (2005) have developed a method to assess the extent to which unobservable factors may be biasing the observed results. Under plausible assumptions, the authors' method permits the calculation

of a point estimate and standard error of the bias resulting from selection on unobservables.[14] Execution of this method returns a negatively signed, but substantively small, bias estimate. Intuitively, this implies that our estimates are unlikely to be significantly biased by unobserved characteristics and the negative sign on the bias implies that, if anything, our estimate of the relationship between baseline MPCP attendance and student achievement growth may be a slight underestimate of the true relationship. More generally, though, the results of this procedure lend confidence to our position that self-selection bias is likely not a significant problem in our study. This is because: (1) the program that we study is targeted to a relatively disadvantaged student population that is more homogenous than a broader student body; and (2) our system of matching up students based on key observable variables, such as neighborhood and baseline test score, likely accounts for some of the unobservable factors that influence both whether a student participates in the MPCP and how well that student subsequently performs on standardized tests.

CAVEATS

The results presented in the preceding sections are limited in their explanatory power in several important ways. Most importantly, the introduction of the test-based accountability policy for MPCP schools in 2010-11 means that this study is no longer solely evaluating the effectiveness of MPCP in promoting student achievement growth. Instead, it is evaluating the effectiveness of both MPCP and the accountability policy that was introduced in 2010-11, and there is some evidence that the larger achievement growth we observe of the students in our MPCP sample is attributable to the introduction of the accountability policy, rather than solely to the effectiveness of MPCP. It is important to stress that, although we have uncovered evidence suggesting that the final-year surge in achievement growth for the MPCP students in our study was due to the new accountability system, we cannot say for sure how the accountability system increased the scores. It is possible that the MPCP schools did a more effective job teaching their MPCP students in the final year of our study because the individual schools' aggregate test scores would be made public. It is also possible that the schools simply took the WKCE testing more seriously in the final year of our study, when the test score results held high stakes for the schools, compared to in the previous years of our study, when the scores were used anonymously and only for research purposes. Test-based accountability systems are supposed to work to boost test scores in the former way but, when first introduced, may actually work in the latter way.

---

14   Specifically, the method for estimating the bias is valid under the condition that selection on unobservables is equal to selection on observables. Slightly more formally, it is valid under the condition that the covariance of the treatment and the mean of the distribution of the index of observables is the same as the covariance of the treatment and the mean of the distribution of the index of unobservables, after adjusting for differences in the variance of the distributions. This condition requires strong assumptions, including 1) that the set of observed variables is chosen at random from the full set of variables that determine voucher receipt and student achievement, and 2) that the number of observed and unobserved variables is large enough that none of the elements dominates the distribution of voucher receipt or student achievement.

Additional caveats are related to data that are missing, 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 had, on average, baseline test scores that were no different from students who remained in the sample. There were, however, some differences by gender and race. Students who remained in the sample were slightly more likely to be female and slightly less likely to be White than students who could not be located. In examining missing students, there were some differences in student characteristics between those missing from the MPCP or the MPS panels. More MPCP students are missing and their baseline math scores are substantially lower than the baseline math scores of missing MPS panelists. In addition, missing students in MPS are slightly more likely to be female than missing MPCP students. There are no differences by 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 baseline characteristics in all our analyses. The fact that we have backfilled missing data regarding permanent demographic characteristics of students, imputed missing data on demographics that we cannot backfill, weighted for missing test scores, and searched for missing students using a number of methods, including telephone surveys and database searches, should help assuage concerns that our results are being affected by issues related to missing data, but readers should be aware of these potential threats to validity.

SUMMARY AND CONCLUSIONS

This report presents the fifth year and final analysis of academic achievement growth 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 mean test scores and other descriptive statistics indicated that in some grades students who attended MPCP schools in 2006 exhibited greater growth in reading achievement from 2006 to 2010 than a group of matched MPS students. This finding was repeated when we estimated multivariate models that included baseline test scores and student demographic variables. Test scores in mathematics also favored MPCP students but the results did not reach statistical significance in most of our statistical comparisons.

These findings come as somewhat of a surprise because, as individuals who have followed this study will know, there were no systematic differences in growth in student achievement across our MPCP and MPS panels between 2006 and 2009, the first four years of this study. However, the finding that the students in the MPCP sample exhibited greater growth in reading achievement should be interpreted with some caution because of the introduction of a test-based accountability policy in the 2010-11 school year. Because the accountability policy was introduced after we carefully matched our random sample of MPCP students to MPS students, this study is no longer solely evaluating the effectiveness of MPCP. Rather, it is evaluating the effectiveness of both MPCP and the accountability policy that was introduced in 2010-11. Our first supplemental analyses presented evidence that the larger achievement growth of the MPCP sample may be attributable to the introduction of the accountability policy, rather than solely due to the greater effectiveness of MPCP.

Other supplemental analyses demonstrate that, conditional on spending any years in MPCP, attending MPCP schools the full five years results in considerably greater levels of achievement growth than attending these schools only one or two years; this result is very consistent with earlier reports. Another one of our supplemental analyses examines the relationship between baseline MPCP attendance and four-year student achievement growth at different points across the achievement distribution. The results suggest that, in mathematics, the effect of baseline MPCP attendance for students at the lower end of the achievement distribution may be somewhat smaller than the effect for students at higher points in the achievement distribution. The results in reading, however, indicate that the relationship between baseline MPCP attendance and achievement growth might actually be somewhat stronger for students at the lower end of the distribution than for students in the 25-90th percentiles.

Five years ago the state of Wisconsin gave us a job: to conduct a five-year longitudinal evaluation of the Milwaukee Parental Choice Program. Across the first four years of that study, the MPCP was relatively stable regarding the accountability policies that applied to the private schools serving the low-income students in the program. During that period we observed no clear differences in achievement growth between our representative sample of MPCP students and a carefully matched sample of MPS students. In the final year of our study, a test-based accountability policy was applied to the MPCP and we observed MPCP students move ahead of their MPS peers in achievement, most clearly in reading. As we have stressed repeatedly, we cannot be certain how much of that one-year achievement gain was solely due to the school choice opportunity presented by the MPCP, how much was solely due to the accountability policy, and how much was due to a symbiosis between the two factors. Our supplemental analyses provide substantial evidence that the accountability policy could be responsible, in large part, for the higher achievement gains of the voucher students.

Our study established a common educational starting point for MPCP and MPS students. The gun sounded and a five-year race was run. The test-based accountability policy was applied to our MPCP runners during the final leg of the race, like a surge of adrenaline, and they clearly crossed the finish line ahead of their MPS peers in reading. That is what we learned from our Longitudinal Educational Growth Study and we think the lesson is an important one.

REFERENCES

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on observed and unobserved variables: assessing the effectiveness of catholic schools." *Journal of Political Economy* 113: 151–184.

Ballou, Dale. 2008. "Test Scaling and Value Added Measurement." Paper presented at the National Conference on Value Added Modeling. Wisconsin Center for Education Research, April 22-24.

Carnoy, Martin & Susanna Loeb. 2002. "Does external accountability affect student outcomes? A crossstate analysis". *Educational Evaluation and Policy Analysis* 24: 305–331.

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. "Going Public: Who Leaves a Large, Longstanding, and Widely Available Urban Voucher Program." *American Educational Research Journal*, 2011" (Forthcoming).

Cowen, Joshua M., David J. Fleming, John F. Witte, and Patrick J. Wolf. 2012. "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

Dee, Thomas S. and Brian Jacob. 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management* 30(3): 418-446.

Hanushek, Eric A. and Margaret E. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management* 24: 297–327.

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.

Jacob, Brian A. 2005. "Accountability, incentives and behavior: Evidence from school reform in Chicago." *Journal of Public Economics* 89: 761–796.

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., Deven Carlson, Joshua M. Cowen, David J. Fleming, and Patrick J. Wolf. 2011. "MPCP Longitudinal Educational Growth Study: Fourth Year Report." SCDP Milwaukee Evaluation Report #23.

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.		Table A-2.	
Descriptive Statistics for Variables Used in Table 4		**Descriptive Statistics for Variables Used in Table 5**	

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

	MPS Matched Counts	MPCP Counts
	N (%)	N (%)
Female	369 (54.7)	353 (55.5)
White	69 (10.3)	49 (7.7)
Black	430 (64.0)	390 (61.3)
Hispanic	147*** (21.9)	182 (28.6)
Asian	19 (2.8)	14 (2.2)
Nat. Am.	7** (1.0)	1 (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 1308 students in the estimation sample for the math achievement models.

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

	MPS Matched Counts	MPCP Counts
	N (%)	N (%)
Female	323 (54.3)	153 (60.2)
White	62 (10.4)	34 (13.4)
Black	370*** (62.2)	112 (44.1)
Hispanic	139*** (23.4)	103 (40.6)
Asian	18 (3.0)	5 (2.0)
Nat. Am.	6 (1.0)	0 (0.0)

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10, based on a two-tailed T-Test. Calculations performed over the 849 students in the estimation sample for the math achievement stayer model.

APPENDIX B – Study Attrition

Of the original 4,007 students in the combined MPS and MPCP panels that we should have been able to locate in the 2010-11 school year, we were unable to locate 1,280 (32 percent) in Year 5. The rate is lower for MPS students (28 percent) compared to students who began our study in the MPCP (36 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. This 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, for students who were in grades 3-8 at baseline (i.e. those that were not scheduled to have graduated from high school before the 2010-11 school year). 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 some racial and gender differences between missing and non-missing students. Specifically, missing students are slightly less likely to be female, but slightly more likely to be White, relative to non-missing students. However,

Table B-1. Sample Attrition Statistics 2006-10

	Non-Missing Students	Missing Students
Average Mean Baseline Math	-0.20	-0.20
Average Mean Baseline Reading	-0.16	-0.11
%Female	53.98*	50.94
%White	8.07***	10.86
%Black	65.31	65.70
%Hispanic	23.36	20.70
%Asian	2.71	2.50
%Native American	0.51*	0.16
%Baseline Grade 3	16.43	18.28
% Baseline Grade 4	16.72	15.00
% Baseline Grade5	15.55***	19.69
% Baseline Grade 6	17.24*	14.84
% Baseline Grade 7	15.51	14.30
%Baseline Grade 8	9.21***	15.48
% Baseline Grade 9	5.68***	NA

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.

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-2 provides evidence on the difference in missing students by sector. Among students we were not able to locate at Year 5, there were no statistically significant differences in mean baseline reading scores between the two sectors. However, baseline math scores were lower for missing MPCP students than missing MPS students. In addition, missing MPCP students were slightly less likely to be female than their missing MPS counterparts. There are no differences in race/ethnicity among the missing students in the two sectors. There are some grade differences, however, as baseline 5th graders make up a smaller proportion of missing MPS students than missing MPCP students while baseline 8th graders make up a slightly larger proportion of missing MPS students. 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). Our multivariate models also include controls for baseline test scores, further mitigating any bias on those variables.

Table B-2. MPS vs. MPCP Attrition Statistics 2006-10

	MPS	MPCP
Missing Students	562 (27.99)***	718 (35.92)
Average Mean Baseline Math	-0.088***	-0.285
Average Mean Baseline Reading	-0.105	-0.118
%Female	302 (53.74)*	350 (48.75)
%White	60 (10.68)	79 (11.00)
%Black	360 (64.06)	481 (66.99)
%Hispanic	125 (22.24)	140 (19.50)
%Asian	17 (3.02)	15 (2.09)
%Native American	0 (0.00)	2 (0.28)
%Baseline Grade 3	107 (19.04)	127 (17.69)
% Baseline Grade 4	86 (15.30)	106 (14.76)
% Baseline Grade5	95 (16.90)**	157 (21.87)
% Baseline Grade 6	75 (13.35)	115 (16.02)
% Baseline Grade 7	87 (15.48)	96 (13.37)
%Baseline Grade 8	112 (19.93)*	117 (16.30)
% Baseline Grade 9	NA	NA

Stars indicate MPS different from MPCP statistics at ***p<0.01, **p<0.05, *p<0.10. Percentages are in parentheses.

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 over time 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 deviation changes from the original scores for each year and none of our estimates remotely approach changes of that magnitude.

Table C-1. Sector Comparisons of 2006 Baseline Scores for Students with WKCE Tests (In Z-Scores): 2006-2010

Subject	MPCP			MPS		
	N	Mean	SD	N	Mean	SD
All Students 2006:						
2006 Math Test***	1927	-0.266	0.990	1926	-0.128	0.980
2006 Reading Test	1927	-0.143	0.983	1926	-0.141	1.000
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
All Students 2010:						
2006 Math Test***	646	-0.364	0.980	672	-0.162	1.046
2006 Reading Test**	643	-0.234	0.944	674	-0.129	1.002

NOTE: *** indicates that baseline scores of MPCP students are different from baseline scores of MPS students at $p<0.01$.
** indicates that baseline scores of MPCP students are different from baseline scores of MPS students at $p<0.05$.

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 an assistant professor at the Martin School of Public Policy and Administration at the University of Kentucky. His research focuses on school choice programs, teacher and student mobility, and organizational issues in schools, and has been funded by the Spencer Foundation, among other entities. He is author or co-author of papers in journals such as the *American Educational Research Journal*, the *American Journal of Education*, *Economics of Education Review* and *Education Finance and Policy*. He received his Ph.D. from the University of Wisconsin-Madison, where he was a graduate fellow in education science, sponsored by the U.S. Department of Education.

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