Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee*

Rajashri Chakrabarti†
Federal Reserve Bank of New York

Abstract

The Milwaukee voucher program, as implemented in 1990, allowed only non-sectarian private schools to participate in the program. Following a Wisconsin Supreme Court ruling, the program saw a major shift and entered into its second phase, when religious private schools were allowed to participate for the first time in 1998. This led to more than a three-fold increase in the number of private schools and almost a four-fold increase in the number of choice students. Moreover, due to some changes in funding provisions, the revenue loss per student from vouchers increased in the second phase of the program. This paper analyzes the impacts of increase in competition brought about by these changes on public school performance in Milwaukee. Using data from 1987 to 2002, and a difference-in-differences estimation strategy in trends, the paper finds that these changes have led to an improvement of the public schools in the second phase of the program as compared to the first phase. The results are robust to alternative samples and specifications, and survive several sensitivity checks including correcting for mean reversion. The findings imply that voucher design matters and choice of parameters in a voucher program is crucial as far as impacts on public school incentives and performance are concerned.

Keywords: Vouchers, Public School Performance, Competition, Response, Mean Reversion
JEL Classifications: H4, I21, I28

*I am grateful to Steve Coate, Ron Ehrenberg and Miguel Urquiola for helpful comments, suggestions and encouragement. I thank Robin Boadway, Julie Cullen, Caroline Hoxby, Brian Jacob, George Jakubson, Paul Peterson and seminar participants at Harvard University, Econometric Society Winter Meetings and American Education Finance Association conference for helpful discussions, the Milwaukee Public Schools and the Wisconsin Department of Public Instruction for providing the data used in this analysis, and the Program on Education Policy and Governance at Harvard University for postdoctoral support. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are my own.

†Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. Email: Rajashri.Chakrabarti@ny.frb.org

1 Introduction

Widespread concerns over the performance of public schools have pushed the issue of public school reform to the forefront of policy debate. The focus of public school reform has been on school choice and accountability, and vouchers are among the most hotly debated instruments of school choice. However, not all voucher programs are alike. They often differ in structure and design and these differences affect public school incentives and responses differently. Therefore, understanding the effects of alternative voucher programs is key to designing an effective voucher policy. This paper contributes in this direction by studying how changes in some crucial policy parameters midway through the implementation of the Milwaukee voucher program affected public school incentives and performance.

The Milwaukee parental choice program (MPCP), as implemented in the 1990-91 school year, made all public school households with income at or below 175% of the poverty line eligible for vouchers to attend private schools. Initially only nonsectarian private schools were allowed to participate in the program. The late 1990s saw two major shifts in the program: (i) Following the Wisconsin Supreme Court ruling, religious private schools participated for the first time in the school year 1998-99. This led to a large increase in the number of private schools participating in the program and the number of public school students lost to the program. (ii) Some changes in the state funding formula led to a discontinuous increase in loss of revenue per student with vouchers for the public schools starting from the school year 1999-2000. These two changes led to a significant increase in voucher competition in the second phase of the program.

This paper analyzes the effects of these shifts on the incentives and responses of the Milwaukee public schools. Specifically, it compares the effect of the program after 1998 to that of the initial 1990 program in terms of public school performance of the treated schools. I designate the period before the Wisconsin Supreme Court ruling (1990-91 through 1997-98) as the first phase of the Milwaukee program, and the period after the Supreme Court ruling (that is, 1998-99 onwards) as the second phase of the program.

Using school-level test score data from Wisconsin and implementing a difference-in-differences estimation strategy in trends, the paper first estimates the program effects in each of the first and second phases of the program. Controlling for potentially confounding pre-program time trends and post-program common shocks, the paper finds that the treated schools exhibited greater improvement in the second phase of the program as compared to the first phase. This finding is robust to alternative strategies, specifications and samples, and continue to hold after controlling for other confounding factors such as mean reversion. The finding has important policy implications. It implies that the choice of parameters in a voucher program is key in determining public school response. While the Milwaukee voucher program in its first phase did not have much of a bite, an increase in competition in the second phase through higher private school participation and an increase in per pupil revenue loss from vouchers led to significant improvement in performance of the treated public schools.

A number of empirical studies have looked at the effect of vouchers on the performance of students who move to private schools with vouchers. Such studies in the context of Milwaukee include Witte et al. (1995), Greene et al. (1996) and Rouse (1998). Comparing choice school students with a random sample of Milwaukee public school students and also low income students, Witte et al. (1995) find no statistically significant evidence of gains of choice students. Arguing that the comparison group in Witte et al. was considerably more advantaged than the treated group and using applicants unsuccessful in the lottery as their control group, Greene et al. (1996) find evidence of gains of the choice students in both reading and math. However, as Rouse (1998) argues these effects may be biased by non-random attrition. Using student fixed effects and taking into account attrition, Rouse (1998) compares scores of students selected to attend choice schools with those of unsuccessful applicants and also a random sample of students from the Milwaukee public schools,—she finds evidence of a positive effect on math scores of choice students but no effect in reading.

Prior studies investigating the effect of vouchers on public schools in the U.S. have mainly looked at the Florida and Milwaukee voucher programs. In the context of Florida, Greene (2001) and Greene and Winters (2003) and West and Peterson (2006) find that the Florida program has led to positive and significant improvements in performance of the treated schools. Figlio and Rouse (2006) find some evidence of improvement of the treated schools in the high stakes tests in Florida, but these effects diminish in the low stakes test.

The present study is most closely related to Hoxby (2003a,b), and it has been greatly informed by these two studies. Hoxby (2003a) analyzes the impact of the Milwaukee voucher program on public schools after the Wisconsin Supreme Court ruling of 1998. Since the MPS students eligible for free or reduced price lunches were the ones eligible for vouchers (see footnote 8), the extent of treatment of the Milwaukee schools depended on the percentages of their students eligible for free or reduced price lunches. Exploiting this, she classifies the Milwaukee schools into two treatment groups ("most treated" and "somewhat treated") based on the percentages of their free or reduced price lunch students. Since all schools in Milwaukee were potentially affected by the program, she chooses, as her control group, a set of schools within Wisconsin but outside Milwaukee that were most similar to the Milwaukee schools. Using a difference-in-differences strategy, she finds a positive productivity response to vouchers. Hoxby (2003b) controls for pre-program differences in trends (unlike Hoxby (2003a)), analyzes post-program data up to 2002[1] (unlike 2000 in Hoxby (2003a)) and using the same treatment-control classification, finds evidence of a positive productivity response to vouchers in Milwaukee after the Wisconsin Supreme Court ruling.

This paper follows Hoxby in the treatment-control group classification. As in Hoxby, it uses the percentages of free or reduced price lunch eligible students in the Milwaukee public schools to classify them into treatment groups. The control group criteria is also based on Hoxby. While it uses and builds upon Hoxby's contribution, it differs from Hoxby (2003a,b) in several important ways. First, the basic question posed is different. The focus of this paper is on voucher design. It is interested in analyzing the effects of changes in some policy parameters in a voucher program,—more specifically, in investigating whether a voucher program characterized by a higher private school participation and higher public school revenue loss per student is able to induce a higher public school performance. Therefore, unlike Hoxby, it compares the effect of the Milwaukee program on public school performance in the second phase with that in the first phase.

Second, while the treatment-control strategy here follows Hoxby (2003a,b), there are some important differences. These are discussed in section 6.1. Third, unlike Hoxby, the current study controls for the potentially confounding factor, mean reversion. Since the more treated schools were also the lowest scoring schools, a potential concern is that any improvement of these schools may be due to regression to the mean rather than a program effect. Fourth, unlike Hoxby, this study controls for the possibility

[1] For the remainder of the paper, I will refer to school years by the calendar year of the spring semester.

that changes in student composition of schools may bias the estimated effects of voucher competition. Fifth, while the graphical analysis in Hoxby (2003b) looks at the effect of the reform for the different years during 1999-2002, the more precise regression analysis looks at the average annual effect of the program (upto 2002 in Hoxby (2003b) and 2000 in Hoxby (2003a)). This paper analyzes the gains of the different treated groups in each of the years separately[2], after the program shift as well as after the initial program. This is instructive since public school response may vary across the different years after program. Finally, unlike Hoxby, this study also investigates whether the potential competition faced by the Milwaukee schools was actually effective. Even if schools had a substantial proportion of their students eligible for vouchers, competition faced by them would not be functional unless there was enough choice school (private schools that accepted voucher students) presence near them. To assess the extent of effective competition, I investigate the distribution of choice schools around Milwaukee public schools and also the extent of the latter's actual loss of students to the program.

Although there are multiple papers that analyze the effects of alternative voucher policies on stratification, distribution and welfare (Epple and Romano (2002), Hoxby (2001), Nechyba (2000), Caucutt (2002)), to the best of my knowledge, there is only one paper so far that looks at the impact of alternative voucher designs on public school performance. Focusing on two publicly funded voucher programs in the U.S.—Florida and Milwaukee—Chakrabarti (2004) shows, both theoretically and empirically, that differences in designs in these two programs have led to very different effects on public school performance in these two places. Specifically, it shows that the "threat of voucher" design in the former has led to a much higher improvement of the treated public schools than the traditional vouchers in the latter. The present study complements this paper in the sense that it shows that changes in some crucial policy parameters even within the same traditional voucher program can have markedly different effects on public school incentives and performance. Given the sparse literature on the effect of voucher design on public school incentives and performance and the policy relevance of this issue, this study makes an important contribution to the literature. The findings imply that voucher design matters and choice of parameters in a voucher program is key as far as effects on public school incentives and response are

[2] This paper focuses on the effect of the program through 2002, as the post-2002 effects are likely to be contaminated by the effects of the federal No Child Left Behind (NCLB) law, more so because the schools treated more by the Milwaukee program are also likely to have been treated more by NCLB.

concerned.

2 Institutional Background and Data

2.1 The Program and its Shifts

The Milwaukee parental choice program (MPCP) was implemented in the 1990-91 school year in the city of Milwaukee. It made all Milwaukee Public School (MPS) students in grades kindergarten through twelve (K-12) with household income at or below 175% of the poverty line eligible for vouchers to attend private schools. Initially, it allowed only nonsectarian private schools to participate to take in voucher students, and student participation in the MPCP was limited to 1% of the MPS membership. Following a Wisconsin Supreme court ruling on June 10, 1998, religious private schools were allowed to participate in the program for the first time in the 1998-99 school year. The participation cap was also raised to 1.5% in 1994-95 and further to 15% in 1996-97.

In spite of the cap on MPCP enrollment, this participation constraint was not binding. As table 1 shows, the number of applicants was almost always less than that allowed by the program. However program growth was limited by the capacity of the participating private schools. The number of private school seats was a binding constraint—as table 1 shows, the number of private school seats was not only well below the number authorized by the statute but was also considerably less than the number of applicants.

Therefore 1998 constituted a benchmark year in the history of the MPCP. The Wisconsin Supreme Court ruling allowing religious schools to participate relaxed the binding constraint of the number of private school seats that limited the growth of the program. Table 2 shows the membership and payment history of the MPCP, and illustrates some of the immediate effects of the ruling. As a consequence of the ruling, the number of private schools participating in the program jumped 3.6 fold from 23 to 83, and the number of students enrolled in the MPCP increased almost four fold from 1497 to 5761. Interestingly, as table 2 shows, MPS membership fell for the first time in the 1998-99 school year.

The financing of the MPCP also changed over the years. Under the MPCP, as implemented in the 1990-91 school year, state aid followed the pupil from the MPS to the private school. Pupils participating in the choice program were included in the membership count for MPS on a prior year basis, even though they were attending private schools under the MPCP. This membership count was then used to calculate the state aid for the district. The voucher amount equaled the state aid per pupil[3] and the MPCP was funded by reducing the state aid for the MPS district by the voucher amount times the number of students attending the MPCP.

Starting from the 1999-2000 school year, the definition of membership was changed to exclude MPCP pupils—unlike earlier, the MPCP pupils were no longer included in the membership count of the MPS for state aid purposes. Moreover, the distribution of the burden of financing of the MPCP was changed. From the 1999-2000 school year, the amount needed to finance the MPCP was funded 50% from a reduction of state aid to the MPS and 50% from a reduction in state aid to the other 425 public school districts in the state.[4]

Although only 50% of the MPCP expenditure came from the MPS from 1999-2000 (45% from 2001-02), the effective loss per student to the MPS was much more in the period since 1999-2000 than before. This was because the membership count of the MPS for state aid purposes no longer included the MPCP pupils, unlike earlier. If v denotes the voucher amount, the loss in state aid per student to the MPS before 2000 was v, while the loss in state aid per student was $(v + 0.5v)$ from 2000.[5] In addition, the voucher amount was larger in the second phase than in the first, which further increased the loss in revenue per pupil in the second phase.

Table 2 also shows the voucher amount, the total MPCP amount and the distribution of the MPCP burden among the MPS and other districts. Note that the relaxation of the private school participation constraint in the 1998-99 school year led to a large increase in the MPCP amount and a consequent reduction in state aid to the MPS. The MPCP amount increased four fold from 7 million in 1997-98 to

[3] More precisely, it was the equalization aid per member. For more details on the institutional background of MPCP, see Chakrabarti(2005).

[4] This was changed from the 2001-02 school year when 45% of the MPCP amount was funded from a reduction in state aid to the MPS and 55% from the state general purpose revenue, so that the other districts did not bear the burden of financing of the MPCP.

[5] Both before 2000, as well as afterwards, MPS had the authority to increase its property tax levy to offset the aid reduction due to MPCP, and to some extent did do so. Note that these increases allowed pertained to the aid reduction on account of MPCP, that is, the MPCP amount before 2000, 50% of the MPCP amount during 2000-2001, and 45% of the MPCP amount from 2002, but not to the loss of revenue due to the inability to count MPCP pupils for state aid purposes. So the exclusion of the MPCP students from the membership count of the MPS still represented a discrete increase in the loss in per pupil revenue in the later period. Also from the perspective of districts and schools, property taxes represent a costlier form of revenue than state aid. Just like the MPS, the other school districts could also increase their property tax levies to offset any aid reductions made due to the MPCP.

28.7 million in 1998-1999 which was funded by a corresponding reduction in state aid to the MPS.[6] The MPCP amount continued to increase due mainly to the increase in the number of choice students, but the 50% funding rule reduced the MPS funding of the MPCP amount. However, it should be noted that this table does not take into account the fact that the membership formula for state aid no longer included the MPCP pupils, (so that the effective loss in state aid from 2000 was much more than illustrated here)—the table only illustrates the distribution of the MPCP burden.

2.2 Data

The data for this paper come from multiple sources and consist of school-level data on test scores, socio-economic characteristics of schools, and school finances. They are obtained from the Wisconsin Department of Public Instruction (DPI), the Milwaukee Public Schools (MPS), and the Common Core of Data (CCD) of the National Center for Education Statistics. Data on socioeconomic characteristics include data on race, sex and percentage of students eligible for free or reduced-price lunches for the period 1987-2002 and are from the CCD and the MPS. Data on per pupil expenditure for the same period are available from the Wisconsin DPI and the MPS.

For the first phase, school-level data on test scores are available on two tests: (i) the Third Grade Reading Test (renamed the Wisconsin Reading Comprehension Test (WRCT) in 1996) obtained from the Wisconsin DPI and (ii) the Iowa Test of Basic Skills (ITBS) obtained from the MPS. The WRCT is a state-administered grade 3 reading test that has been administered since 1989. School scores for this test are available for the three reported "performance standard categories": percentage of students below (% below from now on), percentage of students at (% at), and percentage of students above (% above) the standard for 1989-97. The ITBS reading, math and language arts tests were district administered tests and data on grade 5 ITBS reading, math and language arts scores are available for the periods 1987-93,

[6] Note that some of the MPCP participants were likely to be transfers from private schools (private school students in grades K-3 were eligible to participate). MPS would not have received state funding for these students, and due to the funding formula before 2000, MPS would merely serve as a "pass through entity" for these students' funding, so that the MPCP amount is artificially inflated on this account, as one referee points out. Note that such transfers were not new in the crucial reform year 1998-99, so that the pre-reform MPCP amounts were also inflated by this factor. However, due to the expansion in 1999, these movements were likely to have been larger in 1999. Yet, only a small proportion of the MPCP amount was contributed by movements of students from private schools, —this amount was $2.3 million in 1998-99, which is small compared to the total MPCP amount that year. Also note that this problem of artificial inflation is not attributable to the post-1999 period, as MPCP students were no longer included in MPS membership count for state aid purposes.

1987-97 and 1989-92 respectively.

For the second phase, school level data are available on the Wisconsin Knowledge and Concepts Examination (WKCE) from the DPI. WKCE is a statewide examination administered in grades 4, 8 and 10 annually in the subject areas of reading, language arts, math, science and social studies. The first administration of WKCE in all the three grades took place in 1997. School level grade 4 NPR scores on the five subject areas are available for the period 1997-2002.[7]

The next section analyzes the impact of the two major changes in the program described above on public school performance—the discontinuous increase in private school participation from 1999 and the discrete increase in the loss of revenue per student from 2000, and the effects obtained should be interpreted as a combination of the effects of these two changes. Although the first shift took place a year in advance, responses to shifts and their effects take time to materialize. The two shifts were too close together to separately identify their effects—rather the paper addresses the question as to whether a voucher program characterized by higher private school participation and higher loss in revenue per student is able to bring about higher performance effects from public schools. While the next sections address this question empirically, Appendix A analyzes the effect of these two changes in the context of a theoretical model that captures some of the basic features of the Milwaukee program. It may be noted that the empirical findings are consistent with the predictions of the theoretical model.

3 Empirical Strategy

3.1 Samples and Specifications

With the inception of the program in 1991, all schools in Milwaukee were affected, at least to some extent, as all schools had students eligible for vouchers. Therefore, I first consider all schools in Milwaukee in the pre-program year (1990) as a single treated group, and pick schools outside Milwaukee but within Wisconsin that were as similar as possible to the Milwaukee schools as the control group. I call this sample my "first sample". The control group criteria are discussed below. I restrict my analysis to

[7] Although school scores on WRCT are available for the second phase, the test format as well as the mode of reporting of scores changed. While the pre-1998 reporting categories were % at, % above and % below, the new categories were minimal, basic, proficient and advanced and were not comparable to the earlier ones. Therefore, according to the DPI, a comparable series for WRCT encompassing the two phases is not available. Comparable data for WRCT are available for the second phase only for 1998-2002. Since this makes controlling for pre-program differences in trend impossible (due to availability of only one year of pre-program data), I have not used WRCT scores for my analysis in the second phase.

elementary schools only since there were very few middle and high schools in the MPS, and participation of students in the MPCP was mostly in the elementary grades.

While all Milwaukee schools were treated, there was considerable variation in the extents of treatments of the schools. I use two other strategies of sample formation that exploit this variation.

1. Classification into multiple treatment groups: This strategy is based on Hoxby (2003a,b) and is similar to hers. Since the free or reduced price lunch eligible students of the MPS were the ones eligible for vouchers, the extent of treatment of the Milwaukee schools (or the competition they faced) depended on the percentages of their students eligible for free or reduced price lunches.[8] Exploiting this, Hoxby classifies the Milwaukee schools into two treatment groups based on the percentages of their free or reduced price lunch eligible students in the pre-program period—"most treated" (where at least two-thirds of the students were eligible) and "somewhat treated" (where less than two-thirds were eligible). I classify the schools into three treatment groups (unlike two in Hoxby) based on their pre-program (1990) percentages of free or reduced price lunch students. So the treatment groups here are more homogenous as well as starker from each other. Also, to test the robustness of the results, I consider alternative samples that are obtained by varying the cutoffs that separate the different treatment groups.

Based on the distribution of free or reduced price lunch students in Milwaukee schools, I construct the following samples such that each group contains an appreciable number of schools and the middle group contains the mean and median schools in terms of free or reduced price lunch eligibility. In the 66-47 (60-47) sample, I classify schools that had at least 66% (60%) of their students eligible as "more treated" (MT); schools with eligibility between 66% (60%) and 47% as "somewhat treated" (ST); and schools with less than 47% eligibility as "less treated" (LT). I also consider alternative samples, such as "75-47","66","60","75","50" defined similarly as above (the latter four classify the Milwaukee schools into two treatment groups). The analysis in this paper will use the "66-47" sample, the results for the other samples are similar and are available on request.[9]

[8] While 175% was the cutoff poverty level for eligibility, this cutoff was not strictly enforced (Hoxby (2003b)) and households within this 10% margin (175% and free or reduced price lunch eligibility cutoff 185%) were often allowed to apply. Also there were very few students who fell in the 175%-185% range, in fact 90% of the free or reduced price lunch eligible students qualified for free lunch. (Witte (2000)).

[9] The "66-47" sample contains 33 more treated, 53 somewhat treated and 21 less treated schools. In the more treated group, an average of 84.5% of students were elgible for free or reduced-price lunch, 66.5% were black, and 18.07% were hispanic. In the somewhat treated (less treated) group an average of 55.4% (37.17%) were free or reduced-price lunch eligible, 50.99% (45.37%) were black and 4.09% (3.83%) were hispanic.

The control group criteria are also based on Hoxby (2003a,b), although there is some difference. The control group in Hoxby consists of Wisconsin schools outside Milwaukee that: (i) had at least 25% of their population eligible for free or reduced-price lunch (ii) had black students compose at least 15% of their population, and (iii) were urban. Her control group consists of 12 schools. I designate schools that were located outside Milwaukee but within Wisconsin, satisfied the first two criteria above in 1990 and had locales as similar as possible to the Milwaukee schools in 1990 as my control schools. The locales of the Milwaukee schools fell in two categories,—locales 1 (large central city) and 3 (urban fringe of large central city) as classified by the CCD. No Wisconsin school outside Milwaukee had a locale code of 1. Therefore I picked schools that had locale codes of 2 (middle-size central city), 3 or 4 (urban fringe of mid-size city). Since Wisconsin schools outside Milwaukee were much more advantaged than the Milwaukee schools, I could find only 33 elementary schools that satisfied these characteristics. These schools mostly had locale codes of 2 and were also geographically located close to Milwaukee. In this untreated comparison group, 44.95% of the students were eligible for free or reduced-price lunches, 22.37% were black, and 14.84% were hispanic. This group of schools will serve as my control group for each of the above samples.

Continuous Treatment Variable: A disadvantage of the above strategy is that it constrains the program effect to be the same for all schools within a treatment group. Therefore, an alternative way to assess the impact of the program is to consider a continuous treatment variable. Here the intensity of treatment of schools is proxied by the percentage of their students eligible for free or reduced-price lunches in 1990 ($\%frl$). Still another advantage of this strategy is that it obviates the necessity of the assignment of cutoffs, whose locations may to some extent be debatable.

There was a wide variation across Milwaukee schools in the percentage of their free or reduced-price lunch students. For example, in 1990, while some schools had only 22% of their students eligible, in some others this number was as high as 93%. Exploiting this variation, I investigate whether an increase in the intensity of treatment was associated with higher improvement in each of the first and second phases of the program and how the improvement (if any) compares between the two phases.[10]

[10] Note that the above two methods of sample formation have some advantages over the first method where all Milwaukee schools are considered together as a single treatment group. First they provide an additional way of confirming that the effects obtained are indeed program effects,—they enable one to test whether there is a hierarchy in performance response

This study considers public school scores as the outcome variable. Using pre-program data for each of the two phases (1987-90 for phase I and 1997-98 for phase II), I first look for differences in pre-existing trends between the different groups of schools. If the treated and control groups have similar pre-program trends, I use the following set of specifications to investigate whether the treated groups demonstrate a higher improvement in test scores compared to the control group in the post-program era. If the treated groups demonstrate a differential pre-program trend, in addition to estimating these specifications, I also estimate slightly modified versions of them where I control for their pre-program differences in trends. The specifications below are for the second phase. The first phase specifications are the same except that the years are different.

$$s_{it} = f_i + \alpha_0 t + \alpha_1 v + \sum_I \alpha_{2I}(I * v) + \alpha_3(v * t) + \sum_I \alpha_{4I}(I * v * t) + \alpha_5 X_{it} + \epsilon_{it} \qquad (1)$$

where $I = \{\text{all 1990 Milwaukee schools}\}$ for the first sample, $I \in \{MT, ST, LT\}$ for the multiple treated groups sample and $I = \%frl$ for the continuous treatment variable sample; v is the program dummy, $v = 1$ if year > 1998 and 0 otherwise. The variables v and $v * t$ respectively control for post-program common intercept and trend shifts. The coefficients on the interaction terms $(I*v)$ and $(I*v*t)$ estimate the program effects—$\alpha_{2,I}$, capture the intercept shifts and $\alpha_{4,I}$ the trend shifts. Note that one would expect the effects to have a strict hierarchy—MT effects> ST effects> LT effects. All specifications I describe here are fixed effects regressions. I also estimate OLS counterparts of each of these specifications which also include dummies for the different treatment groups (or $\%frl$ for the continuous treatment variable regressions).

Second, I estimate a completely unrestricted model that includes year dummies to control for common year effects and interactions of post-program year dummies with treated dummies (or $\%frl$) to capture post-program year effects.

$$s_{it} = f_i + \sum_{i=1998}^{2002} \gamma_{0i} D_i + \sum_{i=1999}^{2002} \sum_I \gamma_{1iI}(I * D_i) + \gamma_3 X_{it} + \epsilon_{it} \qquad (2)$$

This specification no longer constrains the post-program year-to-year gains to be equal and allows the

of schools that corresponds to the hierarchy in treatment intensity. For example, if an improvement of all Milwaukee schools is driven by an improvement of less treated schools, one has less confidence that it is a program effect. Second, unlike in the latter two methods, the treatment effect in the case of the first sample will be confounded with the effects of any Milwaukee specific shock in the post-program (or post-reform) period.

program effect to vary across the different years. The coefficients of interest here are γ_{1iI} which capture the post-program year effects.

3.2 Mean Reversion

A potentially confounding issue here is mean reversion since the more treated schools were also the lowest scoring schools in each of the subject areas in both 1990 and 1998. Note that, unlike for the the treatment-control group strategies, mean reversion is not a concern for the continuous treatment variable estimation strategy. To address the issue of mean reversion, I use pre-program data for the corresponding phase. The strategy here is to examine whether schools, that before the program shift were similarly low scoring as the more treated schools in 1998, improved relative to the control schools before the program shift. If they did, then this shift can be attributed to mean reversion as this was before the program shift, and this mean reversion effect can be subtracted from the program effect to get the mean reversion corrected effect.

To implement this strategy, I rank the Milwaukee schools in 1998 on the basis of WKCE NPR scores of each subject, and pick schools in 1997 that had the same rank in that subject as the more treated schools in 1998. I call this group of schools the "low" group. Similarly, I construct subject specific "mid" and "high" groups in 1997 that correspond to the somewhat treated and less treated groups respectively in 1998, and look at improvements of these groups relative to the control group during 1997-98. Table 3 presents the corresponding mean reversion estimates for the second phase. While there is some evidence of mean reversion in math, science and social studies, there is no such evidence in reading and language arts. I also investigate mean reversion in phase I using this method, except that the years now are 1989 and 1990 and schools are ranked on the basis of WRCT and ITBS. There is no evidence of mean reversion in the first phase,—these results are available on request.

3.3 Sorting

Vouchers affect public school quality not only through direct public school response but also through changes in student composition and peer quality brought about by sorting. All these three factors get reflected in the public school scores. (See Hsieh and Urquiola (2006) for a discussion.) So one might argue that any effect(s) observed are driven by student sorting. However, each of the regressions control for socio-economic and demographic composition of schools, so this factor is, at least to some extent, controlled for though any change in student composition in terms of unobservable factors may not be controlled for. Note also that if sorting leads to cream-skimming, this will lead to underestimates of the program effect in each of the two phases, especially in the second phase where the loss of students was higher.

To investigate further this issue, I examine whether the demographic composition of the different Milwaukee treated groups changed after the program shifts. As table 4 shows, there is little evidence of significant changes in demographic compositions of schools, either in phase I or in phase II. This provides suggestive evidence that sorting was not an important factor.

3.4 Competitive Effect–Presence of choice schools

Another concern is whether the competition, at least in the second phase, was effective. As described earlier, a non-negligible proportion of the school population (at least in the more treated ones) were eligible for vouchers. However, the threat of loss of students to the voucher program would not be functional unless there were enough choice schools in close proximity to absorb them. To investigate this issue, I examine the distribution of choice schools in Milwaukee and their distances from public schools. An average Milwaukee school in 2004 had 5.66 choice schools within a one mile radius and 17.23 schools within a 2 mile radius; the corresponding numbers for the more treated schools were respectively 8.13 and 25. As table 5 shows, 27% of the public schools had 1-2 choice schools within a one mile radius, more than 19% had 3-5, 30% had 6-10 and 13% had more than 11 choice schools within a one-mile radius. The presence of choice schools around the more treated schools was even stronger. This picture is mirrored for 2 and 3 mile radii also. Thus there was considerable choice school presence, more so in the vicinity of more treated schools. The more treated schools indeed faced the strongest competition from the program,—not only were a higher proportion of their students eligible for vouchers, but there were also a larger number of choice schools surrounding them.

3.5 Competitive Effect, Intensity of Treatment and Loss of Voucher Students

I also check whether a higher proportion of student eligibility was associated with a corresponding higher loss of voucher students from more treated schools. Table 6 shows the distribution of students lost due to vouchers from more treated, somewhat treated and less treated schools. For ease of comparison, I have normalized the numbers in each year in terms of the students lost by the less treated schools in that year. The losses have the expected hierarchy– the more treated loss always exceeded the somewhat treated loss and the somewhat treated loss the corresponding less treated loss, and these differences were often statistically significant. Thus the table corroborates the fact that higher intensity of treatment was associated with higher loss of voucher students and higher competitive pressure.

4 Results

Using all 1990 Milwaukee schools as the treated group and the control schools, Figure 1 Panel A plots the mean scores for reading, math and language arts for phase I. Panel B does the same for phase II. For all graphs in this paper, the vertical line is the reform bar representing program shift and is drawn between the two years within which the program (or reform) happened. For phase I (phase II) 1990 (1998) is the last pre-program year and 1991 (1999) is the first post-program year. While there is not much evidence of relative improvement in phase I, there seems to have been an improvement of the treated schools relative to the control group in phase II in each of reading, math and language arts. It would be instructive to see whether these patterns hold up in a more rigorous regression analysis.

Using all treated schools as a single treated group, table 7 panel A looks at the effect of the voucher program on treatment status in phase I, while panel B looks at the effect in phase II.[11] Most of the regressions reported here as well as in the other tables in this paper are from fixed effects regressions. Results from the corresponding OLS regressions are similar and hence skipped,—they are available on request. The results for WRCT (both % above and % below) show statistically significant improvements in the second and fourth years. Although most of the other coefficients are positive (negative for % below), they are not statistically significant. In ITBS reading, math and language arts, there is not much evidence of improvement,—while many of the effects are positive, they are not statistically significant.

[11] In phase I, there is no evidence of differences in pre-program trends between treated and control groups in the various subject areas. Same is true for the multiple treatment group and continuous variable samples. In phase II, while there is no evidence of differences in pre-existing trends in reading, language arts and science for the various samples, the treated group in the single treatment group analysis and the more treated group in the multiple treatment group analysis exhibit positive significant trends in comparison to the control group of schools. These results are available on request. Whenever there are differences in pre-program trends, the results reported control for these differences. Also for each of the subject areas, I estimated regressions that control for pre-program trends as well as those that do not. The results are qualitatively similar under both formulations.

Moreover, an F test of joint significance of the treated effects can not reject the hypothesis that the effects are jointly zero in reading and math while in language arts it shows some evidence of deterioration.

The phase II results presented in panel B show a different picture. Results from both models show positive treatment effects which are statistically significant in most cases. Moreover, the F-test for joint significance of the treatment effects always strongly rejects the hypothesis of zero effects. Figure 2 graphs the predicted values for treated and control groups from OLS estimation of model 1. Panel A corresponds to phase I and panel II to phase II. While there is not much evidence of relative improvement in phase I, the treated group shows improvement relative to the control schools in all the three subject areas in phase II. To summarize, the above analysis suggests that the shifts in the voucher program led to improvements of the Milwaukee schools in phase II. However, since the above analysis pooled all treatment schools together in one treatment group, the results do not rule out the possibility that the effects were driven not by schools that were most treated by the program, but by ones that were less treated. If this was indeed the case, then one has less confidence that these effects are driven by the program shifts. A way of confirming that the above effects indeed are program effects is to check whether the treatment effects increased with treatment intensity. The following analysis with multiple treatment groups and continuous treatment intensity variable aims to do that.

Using the 66-47 sample, tables 8 and 9 analyze the effect of the Milwaukee voucher program in phases I and II respectively. For each set, the first column reports results from model 1 and the second from model 2. In phase I, once again there are positive and significant effects in the second and fourth years after program in WRCT. Moreover, they have the right hierarchy in that the more treated effects exceed the corresponding somewhat treated effects and the somewhat treated effects the corresponding less treated effects, though they are not statistically different between groups.[12] Although many of the other effects are positive, they are not significant and do not always have the right hierarchy between groups. There is not much evidence of improvement either in ITBS reading or math. While most of the effects are positive, they are not statistically significant in most cases and do not have the right hierarchy. In language arts, both models show a deterioration immediately after the program, although it is reversed,

[12] The less treated effects do not add any new insight and hence are skipped for lack of space. They are available on request.

to some extent, in the second year after the program. Once again, the somewhat treated effects, exceed the more treated effects.[13] The F-test for joint significance of the more treated effects show no evidence of any effect for ITBS reading and math, and a deterioration in ITBS language arts.

The results for phase II (table 9) present a sharp contrast to those in phase I. In reading and language arts, estimates from both models show positive effects that are statistically significant in most cases. Moreover, the effects almost always have the right hierarchy and are often statstically different between groups. In science, while the coefficients are positive significant in most cases, the somewhat treated effects exceed the more treated effects in the initial years of the second phase. However, the more treated schools have a higher post-reform growth and surpass the somewhat treated schools in the fourth year after program. In math and social studies, estimates from both models show positive and statistically significant effects, but they often do not have the expected hierarchy,—in fact, in both subject areas the somewhat treated effects exceed the corresponding more treated effects economically (though never statistically).

Figure 3 graphs the predicted values for the various treated groups and the control group from OLS estimation of model 1. Panel A presents the graphs for phase I, panel II for phase II. Consistent with the results above, there is not much evidence of program effects in ITBS reading and math in phase I, and some evidence of a deterioration in language arts in the first year. In contrast, the phase II graphs show considerable improvement in reading and language arts after 1998. The more treated group showed the largest improvement (followed by the somewhat treated group) and the gaps between the more treated trend line and those of the other groups have narrowed. In math, as seen in the regression results above, somewhat treated group seems to have improved to a greater extent than the more treated group.

Tables 10 and 11 present the results from the continuous treatment variable strategy, for phases I and II respectively. In phase I, once again the second and fourth year effects in WRCT provide some evidence of improvement economically, although they are not statistically significant. Moreover, there is no evidence of any effect in ITBS reading or math, although once again there is some evidence of deterioration in language arts in the first year. Moreover, in neither of the regressions are the treatment

[13] Since the ITBS was administered in Milwaukee as a district assessment program, I do not have data on non-Milwaukee Wisconsin schools for this test. As a result, my comparison group for the ITBS is the less treated group of schools. Since the comparison group is also treated to some extent, I expect my estimates for the ITBS to be underestimates.

effects jointly significant (except in language arts where there is some evidence of deterioration). The results for phase II (table 11) show considerable evidence of improvement in reading and language arts, at least in the second, third and fourth years after program. There is statistically significant evidence of improvement in science in the second year after program,—the other year effects for science are also positive, although they are not significant. In math and social studies, many of the effects are positive, although they are not statistically significant.

The findings above can be summarized as follows: For phase I, the results are mixed. Most of the coefficients are positive,—however they do not always have the right hierarchy and are often not statistically different from zero. In phase II, there is considerable evidence that the program shifts led to improvements in reading and language arts and these effects increased with the increase in treatment intensity. In science, the initial improvement of the somewhat treated group exceeded that of the more treated group, however over the years the more treated group improved at a higher rate than the somewhat treated group, so much so that the fourth year effect surpassed the corresponding effect for the somewhat treated group. In math and social studies, on the other hand, although most of the effects are positive and often significant, they do not often have the expected hierarchy. The results are robust in that they hold for different samples,[14] different specifications and both OLS and FE estimates for each specification.

Finally, it might be useful to compare the results obtained above for the multiple treatment group analysis with the corresponding effects in Hoxby. In Hoxby (2003b)[15], the average annual effect of being most treated on WKCE language, math and science NPR scores respectively are 7.959, 8.062 and 13.837 respectively. Comparison of these results with the corresponding more treated effects (table 9) shows that the effects are qualitatively similar. The differences in the actual magnitudes can be attributed to differences in methodology and sample,—I consider effects separately over the years unlike average annual effects in Hoxby, the control group here is somewhat different, and the regressions here also control for demographic and socioeconomic characteristics, unlike Hoxby.

[14] In addition to the above three samples (all treated schools and control schools, multiple treated groups and the control schools for the 66-47 sample, continuous treatment variable sample), qualitatively similar results are obtained from the 60-47, 75-47, 66, 60, 75 and 50 samples. These results are available on request.

[15] I consider Hoxby (2003b) because similar to my phase II analysis, the post-reform period considered by the study is 1999-2002. I consider the results for her more treated group only, because her somewhat treated group does not directly correspond to any of my treatment groups in the 66-47 sample.

Milwaukee Phase I Versus Phase II

Table 12 compares the effect of the voucher program in Milwaukee phases I and II. The estimates here are based on estimates from model 2 in tables 7-11 and all figures are expressed in terms of the respective sample standard deviations. (The results from model 1 are similar and hence not reported here.) Since the tests in phase I and II are different, a potential concern here is whether the difference in treatment effects are driven by the differences in tests rather than shifts in the Milwaukee program. First, because of the differences in tests, all treatment effect comparisons between the two phases are made in standard deviation terms, I do not attempt to compare them otherwise. Second, since I use a difference-in-differences estimation strategy, any inherent differences in the tests would be differenced out, as long as they are common shocks. Also, since I have pre-reform test score data for each phase, I am able to control for any pre-existing differences in trends, so that if there are inherent differences in phases I and II tests that affect the treated and control groups differently, these are likely to have been controlled for.

Table 12 panel A considers all treated schools as a single group; panel B considers more treated effects from the multiple treatment group analysis; panel C presents results after correcting the panel B effects for mean reversion; panel D presents results for the continuous treatment variable analysis. The results are consistent across the four panels. In reading, the WKCE phase II effects in each of the years exceed the corresponding phase I effects in WRCT (% above), WRCT (% below) and ITBS, except in fourth year WRCT (% above) for the treatment group analysis. Moreover, while the phase II effects are always statistically different from zero (except in the first year), the phase I effects are often not significant. In language arts, there is no evidence of improvement in phase I. On the other hand, there is positive significant effects in phase II each of which exceed the corresponding phase I effect. In math, the Milwaukee phase I effects are never significant and are often negative. On the other hand, the phase II effects are positive, statistically different from zero in most cases—and each of the phase II effects exceed the corresponding phase I effect. However, it should be noted here, that as seen earlier the effects in math do not have the right hierarchy, the somewhat treated effects in each of the years exceed economically (though, not statistically) the corresponding more treated effects. To summarize, it can be said that the improvement of the more treated schools in Milwaukee phase II has been considerably larger than those in Milwaukee phase I, at least for reading and language arts, and there is no evidence to the contrary in math.[16] This finding is quite robust since it holds for all the different samples, different specifications and different tests.

Consistent with the above findings, there is quite some anecdotal evidence that suggests that schools in Milwaukee have responded to the program in the second phase. In 1995, MPS had one school with before and after-school program. In 2000, there were eighty two such programs. Two MPS schools had health clinics in 1995, in 2000 this number was forty seven. A contract settled between the MPS and the Milwaukee Teachers' Education Association (MTEA) allowed the public schools to hire teachers on the basis of merit, rather than seniority in the second phase. Traditionally, teachers were hired on the basis of seniority only. A teacher-evaluation system was established that had union members weeding out bad teachers.[17]

4.1 Some other Issues

Charter Schools

Milwaukee has seen a recent spurt in the growth of charter schools. Therefore a natural concern is whether the program effect, especially in phase II, is contaminated by a competitive effect from charters. Although charter schools have been allowed to enter in Milwaukee from the 1993-94 school year, the growth of charter schools was initially slow with just 1 charter school till 1998, 3 schools and 186 students in 1999, and 5 schools and 1,239 students in 2000. The charter program picked up in the year 2000-01 with 11 schools enrolling 5,022 students and further to 24 schools and 9,442 students in 2002.

Several points may be noted in this context. First, the charter schools were not a major factor in the first two years of the second phase (1999 and 2000), yet there was considerable improvement of the treated schools. If charter competition was the driving force, there should not have been an improvement in 1999 and 2000. Second, the charter schools became more prominent in school year 2000-2001, however

[16] I also do a pair-wise non parametric test (sign test) for each of panel A and panel B effects, where I ignore the significance of coefficients and consider only their signs. Under the null of equal effects the probability that any one effect size in Milwaukee phase II exceeds the corresponding one in Milwaukee phase I is $\frac{1}{2}$. I have a total of 68 pair-wise comparisons in this table. Under the null, $D =$(Milwaukee phase II effect-Milwaukee phase I effect) follows a binomial $(68, 0.5)$ distribution. D is positive in 65 cases. The probability of getting 65 positive D under the null is $(0.5)^{65}$. Since this is very small, the null of equal effects can be comfortably rejected.

[17] See Hess (2002) and the introduction by Howard Fuller in Carol Innerst (2000).

there is no evidence of any shift in the program effect in 2001 which casts further doubt that the program effects are contaminated by a charter effect. Third, the results remain very similar after dropping 2001 and 2002. Fourth, the identification strategy in the paper is designed to capture the effect of vouchers and to rule out the effects of other potentially confounding factors such as charters. The charter school program was not specific to Milwaukee but was common to entire Wisconsin, so the effect of charter competition would be differenced out, at least to some extent, by the difference-in-differences analysis. What is perhaps more important is that while the treatment intensity from vouchers increased with the Milwaukee schools' proportion of low income students, the treatment intensity from charters did not. This rules out charters as a confounding factor in the strategies that exploit the differences in treatment intensity from vouchers. Since charters were open to all students and were not restricted to low income students only, the more treated schools in Milwaukee were not differentially affected by the charter program. Rather charter competition was a common effect that was faced by all Milwaukee schools. This is further supported by the fact that the distribution of charters around more treated schools was similar to those of an average Milwaukee school. An average Milwaukee school had 2.45 charter schools within a one-mile radius in 2004 while a more treated school had 2.70 schools and the difference is not statistically significant. Since the continuous treatment variable analysis uses only the Milwaukee public schools, the charter effect would be absorbed in the common year effect in this analysis. Also note that the inclusion of year by Milwaukee dummy interactions (and correspondingly excluding year by less treated interactions) in the unrestricted regressions for the multiple treatment group analysis do not change results qualitatively.[18]

Accountability System

Wisconsin also had an accountability system. However, the rules of the accountability system were symmetric for all schools, so that all schools were similarly affected. Therefore, any effect of the accountability system would be absorbed by the year dummies in the non-linear specification and the common intercept and trend shifts in the linear model.Also, the analysis above controls for any differences in

[18] Milwaukee had two other choice programs,—the Chapter 220 Program and PAVE (Partners Advancing Values in Education), established in 1978 and 1993 respectively. Note though that the treatment effects above are not likely to be contaminated by these. Controlling for differences in pre-program trends controls for any effect due to Chapter 220. On the other hand, since PAVE covered only one-half of the private school tuition, overwhelmingly white and more advantaged households participated in the PAVE. In contrast, the more-treated schools were predominantly black and hence were not likely to be strongly affected by PAVE. For a more detailed discussion, see Chakrabarti (2005).

pre-existing trends. So, if the accountability system did affect the different groups differently, this would be controlled for by controlling for the differences in pre-program trends. Thus, this factor is unlikely to bias the program effect.

5 Conclusion

This paper analyzes the role of vouchers as instruments of public school reform and investigates the effect of two major shifts in the Milwaukee voucher program in the late 1990s on public school performance. It finds robust evidence that voucher design matters, and the findings imply that judicious choice of some of the underlying policy parameters in a simple means-tested voucher program can go a long way in inducing public school improvement.

The growth of the Milwaukee voucher program in its initial years was severely limited by the lack of availability of adequate number of private school seats. The number of choice applicants exceeded by far the capacity of the private schools participating in the voucher program. Following a 1998 Wisconsin Supreme court ruling religious schools were allowed to participate for the first time in the 1998-99 school year. This led to a major expansion of the program both in terms of participating choice schools and choice students, and the program entered into its second phase. Due to some changes in the funding formula, the second phase was also characterized by a discontinuous rise in the revenue loss per student from vouchers. These changes led to a significant increase in competition in the second phase of the program.

Using a difference-in-differences analysis in trends and Wisconsin data from 1987 through 2002, the paper shows that these shifts led to a much larger improvement in the second phase compared to the first phase. This result is robust to alternative samples and specifications, and survive several sensitivity checks. The findings imply that any voucher program may not have a positive effect on public school incentives and performance. However, careful choice of parameters can lead to improvement in performance of public schools. The findings of the paper have important implications for public school reform, more so in the context of the present concern over public school performance.

Appendix A: Theoretical Framework

In this appendix, I set up a simple model that captures the basic features of the Milwaukee program, and analyze the effects of the two shifts in this framework. There are three agents in the model: (i) the public school, (ii) the private schools, and (iii) the households. The public school is free and offers quality $q = q(e, b)$, where e and b are public school effort and peer-group quality respectively. The objective of the public school is to maximize net revenue $R(.) = p.N - [c_1 + c(N) + C(e)]$, where p is exogenously given per pupil revenue, c_1 is a fixed cost, N is the number of public school students, e is public school effort, and $c(.)$ and $C(.)$ functions are increasing and strictly convex. There is a continuum of private schools providing a continuum of quality levels. Households pay a tuition $T = t \cdot Q$ ($t > 0$) to attend a private school of quality Q.

Households are characterized by an income-ability tuple (y, α), where $y \in [0, 1]$ and $\alpha \in [0, 1]$; y and α are assumed to be independently and uniformly distributed. The household utility function is given by $U(x, \theta, \alpha) = h(x) + \alpha u(\theta)$, where $h_x, u_\theta > 0, u_{\theta\theta} < 0$, x denotes the numeraire good and θ denotes school quality (public or private). To simplify analysis, I assume $h_{xx} = 0$. If a public school household decides to switch to a private school with vouchers, it incurs a positive switching or relocation cost c.

I model two alternative scenarios: (i) a simple public-private system (PP) without vouchers (the baseline), which can be thought of as the pre-program scenario; and (ii) a Milwaukee-type voucher system. The PP system can be thought of as a two-stage game where the public school chooses effort in the first stage and households choose between schools in the second. The Milwaukee system can be thought of as a three-stage game. The government announces voucher v and a cutoff (or target) income level y_T in the first stage. Only households with $y \leq y_T$ are eligible for vouchers. The public school chooses effort in the second stage and households choose between schools in the third.

The number of private school seats available for households applying with vouchers is limited. Private schools pick voucher students randomly[19], so that a random sample of those that apply are selected,—a certain proportion (say, β) of the applicants are successful and $(1 - \beta)$ proportion are unsuccessful and return to the public school.

[19] This is in keeping with the feature of the Milwaukee program, where private schools are required to pick students randomly if oversubscribed and to accept all students otherwise.

The public-private equilibrium is characterized by an effort-peer quality tuple (e_{PP}, b_{PP}), where (i) e_{PP} is an equilibrium of the stage 1 game, given b_{PP} and (ii) b_{PP} is an equilibrium of the stage 2 game, given e_{PP}. The voucher equilibrium is a peer-group quality b_V and an effort e_V such that given voucher v, income cutoff y_T and proportion β, (i) e_V characterizes the public school equilibrium, given b_V and (ii) b_V characterizes the household equilibrium, given e_V.

Household behavior: Suppose all households expect a peer group quality $b^e \in [0,1]$. Then for each y and given t, v, e, c, b^e, there exists a unique household $0 < \hat{\alpha} < 1$ such that all households with lower ability choose the public school and those with higher ability choose a private school. This $\hat{\alpha}$ is the unique solution to the equation: $[h(y + v - t.Q^* - c) + \alpha u(Q^*)] - [h(y) + \alpha u(q(e, b^e))] = 0$ (A.1) where Q^* is the optimal private school quality choice of the household $(y, \hat{\alpha}(y))$, and v takes on a value of 0 under the public-private system, and an exogenously given positive value under the Milwaukee system iff $y \leq y_T$. Given b^e, peer group quality b is given by:

$$b = \frac{\int_0^{y_T}\int_0^{\hat{\alpha}(y,b^e,v,.)} \alpha d\alpha dy + (1-\beta)\int_0^{y_T}\int_{\hat{\alpha}(y,b^e,v,.)}^{\hat{\alpha}_0(y)} \alpha d\alpha dy + \int_{y_T}^1\int_0^{\hat{\alpha}(y,b^e,0,.)} \alpha d\alpha dy}{\int_0^{y_T}\int_0^{\hat{\alpha}(y,b^e,v,.)} d\alpha dy + (1-\beta)\int_0^{y_T}\int_{\hat{\alpha}(y,b^e,v,.)}^{\hat{\alpha}_0(y)} d\alpha dy + \int_{y_T}^1\int_0^{\hat{\alpha}(y,b^e,0,.)} d\alpha dy} = g(b^e, e, v, t, c, y_T, \beta) \quad \text{(A.2)}$$

At equilibrium, b corroborates the initial conjecture b^e, that is, $b = b^e$. (A.3)

The household equilibrium is characterized by (A.1)-(A.3) and the equilibrium peer quality satisfies $b^* = g(b^*, e, v, t, c, y_T, \beta)$. The equilibrium number of public school students increases with public school effort and decreases with vouchers. An increase in the proportion β leads to a decrease in the number of public school students at a household equilibrium, but an increase in the marginal number of students that the public school can gain with an increase in effort.[20]

Public School Behavior: The public school correctly anticipates household behavior in the future stage of the corresponding game, and chooses effort to maximize net revenue. Based on the funding rules, the net revenue functions in the two phases can be written as: $R_{V,I}(.) = pN_0 - v[N_0 - N(e, v)]$ $-c_1 - c(N(e, v)) - C(e)$, where $N_0 = N(e^*, 0)$ gives the equilibrium number of students under the pre-program public-private system, and $R_{V,II}(.) = pN(e, v) - \frac{v}{2}[N_0 - N(e, v)] - c_1 - c(N(e, v)) - C(e)$ $= pN_0 - (p + \frac{v}{2})[N_0 - N(e, v)] - c_1 - c(N(e, v)) - C(e)$. Since $(p + \frac{v}{2}) > v$, loss in revenue per student (l) in phase II exceeds that in phase I.

[20] For proof of existence of equilibrium as well as intuitions and formal proofs of the statements in this paragraph and the proposition below, see Chakrabarti (2005). Note that $\hat{\alpha}_0(y)$ in $A.2$ is the cutoff ability under the pre-program public-private system and is independent of any parameter change in the voucher system.

Proposition 1 *(i) In a voucher program, an increase in the revenue loss per student due to vouchers, l, increases equilibrium effort. (ii) In a voucher program, an increase in the proportion β increases equilibrium effort.*

Note that both l and β were higher in the second phase. An increase in l implies that the revenue that can be gained by attracting a student is higher. This induces public schools to supply a higher effort at equilibrium in phase II. Consider the second part of the proposition. Suppose there are two voucher systems, the only difference between them being a higher β in the second one. At equilibrium, under the first system, marginal revenue of effort equals its marginal cost. Starting from this effort in the second system, a marginal increase in effort attracts more students than in the first system, while marginal cost is actually lower. Therefore the public school finds it optimal to supply a higher effort under the second system. Using this result and the definition of quality and peer quality, the corollary below follows.

Corollary 1 *Equilibrium effort and quality under Milwaukee Phase II will exceed those under Phase I.*

References

Caucutt, Elizabeth (2002), "Educational Vouchers when there are Peer Group Effects-Size Matters" International Economic Review, 43(1), 195-222.

Chakrabarti, Rajashri (2004),"Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs," Working Paper, Harvard University.

Chakrabarti, Rajashri (2005),"Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee," Working Paper, Harvard University.

Epple, Dennis and Richard Romano (2002), "Educational Vouchers and Cream Skimming" NBER Working Paper # 9354.

Figlio, David and Cecilia Rouse (2004), "Do Accountability and Voucher Threats Improve Low-Performing Schools?", Journal of Public Economics, 90, 2006, 239-255.

Greene, Jay, Paul Peterson, Jiangtao Du, Leesa Boeger and Curtis Frazier (1996), "The effectiveness of School Choice in Milwaukee: A secondary Analysis of Data from the Program's Evaluation," mimeo, University of Houston.

Greene, Jay (2001), "An Evaluation of the Florida A-Plus Accountability and School Choice Program," New York: Manhattan Institute for Policy Research.

Greene, Jay and Marcus Winters (2003), "When Schools Compete: The Effects of Vouchers on Florida Public School Achievement," Education Working Paper 2.

Hoxby, Caroline (2001), "Ideal Vouchers", Harvard University typescript.

Hoxby, Caroline (2003a), "School Choice and School Productivity (Or, Could School Choice be the tide that lifts all boats?)," in C. Hoxby ed. The Economics of School Choice, University of Chicago Press.

Hoxby, Caroline (2003b), "School Choice and School Competition: Evidence from the United States," Swedish Economic Policy Review.

Hess, Frederick (2002), "Revolution at the Margins: The Impact of Competition on Urban School Systems" Brookings Institution Press, Washington, D.C.

Hsieh Chang-Tai and Miguel Urquiola (2006), "The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program," Journal of Public Economics, 90, 1477-1503.

Innerst, Carol (2000), "Competing to win: How Florida's A-Plan has Triggered Public School Reform," The Center for education Reform.

Nechyba, Thomas J. (2000), "Mobility, Targeting, and Private-School Vouchers", American Economic Review 90(1), 130-146.

Rouse, Cecilia (1998), "Private School Vouchers and Student Achievement: Evidence from the Milwaukee Choice Program," Quarterly Journal of Economics 113(2), 553-602.

West, Martin and Paul Peterson (2005), "The Efficacy of Choice Threats within School Accountability Systems: Results from Legislatively Induced Experiments," The Economic Journal, 116, Issue 510: C46-C62.

Witte, John, Troy Sterr and Christopher Thorn (1995), "Fifth-year Report: Milwaukee Parental Choice Program," mimeo, University of Wisconsin.

Witte, John (2000), "The Market Approach to Education: An analysis of America's First Voucher Program," Princeton University Press, Princeton.

Table 1: Milwaukee Parental Choice Program Participation

	School Year				
	1990-91	1991-92	1992-93	1993-94	1994-95
Number of Students Allowed by Statute	931	946	950	968	1,452
Number of Private Nonsectarian Schools in Milwaukee	22	22	23	23	23
Number of Private Schools Participating	7	6	11	12	12
Number of Seats Offered in Private Schools	406	561	694	811	982
Number of Students who Applied	577	689	998	1049	1046

Source: Wisconsin Legislative Audit Bureau Report 95-3 (1995).

Table 2: Milwaukee Parental Choice Program: Membership and Payment History

						Funding of the MPCP Amount			
						MPS		All Other Districts	
Year	Number * of Schools	Aid ** Members	MPS Enrollment	Voucher	MPCP Amount (Millions)	Reduction (Millions)	% of Aid	Reduction (Millions)	% of Each District's Aid
1990-91	7	300		$2446	$0.7	$0.7	0.3	$0.0	0.0
1991-92	6	512	93,381	2643	1.4	1.4	0.5	0.0	0.0
1992-93	11	594	94,258	2745	1.6	1.6	0.6	0.0	0.0
1993-94	12	704	95,258	2985	2.1	2.1	0.7	0.0	0.0
1994-95	12	771	98,009	3209	2.5	2.5	0.8	0.0	0.0
1995-96	17	1288	98,378	3667	4.6	4.6	1.2	0.0	0.0
1996-97	20	1616	101,007	4373	7.1	7.1	1.6	0.0	0.0
1997-98	23	1497	101,253	4696	7.0	7.0	1.5	0.0	0.0
1998-99	**83**	**5761**	**99,814**	**4894**	**28.7**	**28.7**	**5.6**	**0.0**	**0.0**
1999-00	90	7575	99,729	5106	39.1	19.5	3.4	19.5	0.6
2000-01	100	9238	97,985	5326	49.0	24.5	4.1	24.5	0.7
2001-02	102	10497	97,762	5553	59.4	26.7	4.4	0.0	0.0
2002-03	102	11350	97,293	5783	65.6	29.5	4.7	0.0	0.0

* Represents the number of choice schools.

**Aid membership is calculated as the average of September and January FTE, plus summer school.

Sources: Wisconsin Legislative Fiscal Bureau Informational Paper 29 (2003) and Wisconsin Department of Public Instruction.

Table 3: Mean Reversion in Wisconsin, 1997-1998

(Using WKCE Reading, Language Arts, Math, Science and Social Studies Scores, 1997-98)

	Based on Individual Subject Score Rank									
	Reading		Language Arts		Math		Science		Social Studies	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Low * trend	1.63	2.20	-1.09	-1.65	1.25	3.43*	2.42	4.29**	3.06	4.70**
	(3.52)	(2.51)	(3.14)	(2.06)	(3.63)	(1.87)	(3.27)	(1.86)	(3.03)	(2.26)
Mid * trend	-2.80	-3.06	-3.71	-3.13	1.83	4.23**	-0.10	-0.63	0.77	0.91
	(2.91)	(1.97)	(2.79)	(1.85)	(3.15)	(1.90)	(2.88)	(1.68)	(2.31)	(1.56)
High * trend	-3.56	-2.54	-6.00	-5.85	1.77	2.07	-3.21	-4.39	-4.20	-2.93
	(3.71)	(2.73)	(4.10)	(2.73)	(5.39)	(3.10)	(4.79)	(2.45)	(4.18)	(2.56)
Observations	229	229	229	229	230	230	230	230	230	230
R^2	0.55	0.91	0.53	0.92	0.62	0.93	0.66	0.94	0.68	0.93

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. Huber-White standard errors are in parenthesis. OLS regressions include dummies for low, mid and high groups respectively while fixed effects columns include school fixed effects. All regressions are weighted by the number of students tested and include race, sex, free or reduced price lunch eligibility and per pupil expenditure as controls. Standard deviation of WKCE Reading scores = 12.86, Standard deviation of WKCE Language Arts scores = 12.16, Standard deviation of WKCE Math scores = 14.18. Standard deviation of WKCE Science scores = 13.94, Standard deviation of WKCE Social Studies scores = 12.83.

Table 4: Effect of the Milwaukee Program on Demographic Composition of Schools

	Phase I			Phase II		
	% Black	% Hispanic	% White	% Black	% Hispanic	% White
Less treated * program	0.90	0.40	-1.26	1.58	-0.97	-0.84
	(1.59)	(0.83)	(1.38)	(1.97)	(2.17)	(1.25)
Somewhat treated * program	-0.25	1.06	-1.24	1.80*	0.30	-2.38***
	(1.35)	(0.63)	(1.16)	(1.04)	(0.80)	(0.89)
More treated * program	-1.0	1.57	-1.24	0.42	0.28	-0.42
	(1.34)	(0.81)	(1.09)	(0.90)	(0.72)	(0.75)
Less treated * program * trend	0.22	0.16	-0.69***	-1.46	0.43	0.89*
	(0.32)	(0.15)	(0.27)	(0.90)	(1.12)	(0.51)
Somewhat treated * program * trend	0.70	-0.12	-0.89***	-1.21***	-0.02	1.06***
	(0.25)	(0.13)	(0.20)	(0.39)	(0.32)	(0.32)
More treated * program * trend	0.08	-0.39***	0.61***	-0.29	-0.80***	1.38***
	(0.23)	(0.14)	(0.19)	(0.33)	(0.27)	(0.25)
Observations	1228	1226	1228	771	771	771
R-squared	0.95	0.97	0.97	0.99	0.97	0.98

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. This table uses the 66-47 sample. Huber-White standard errors are in parenthesis. All regressions include school fixed effects, time trend, program dummy and program dummy interacted with trend.

Table 5: Distribution of Private Schools Within 1, 2, 3 Mile Radii of Public Schools

	Number of Private Schools Within 1 Mile Radius					
	0	1-2	3-5	6-10	11-15	>15
% of Public Schools	10.68	27.18	19.42	30.1	11.65	0.97
% of More Treated Public Schools	3.22	0	22.58	48.38	22.58	3.22
	Number of Private Schools Within 2 Mile Radius					
	0	1-5	6-10	11-20	21-30	>30
% of Public Schools	0.97	17.48	12.62	31.07	25.24	12.62
% of More Treated Public Schools	0	3.22	0	25.81	45.16	25.81
	Number of Private Schools Within 3 Mile Radius					
	0	1-10	11-20	21-30	31-40	>40
% of Public Schools	0	14.56	14.56	16.5	17.48	36.89
% of More Treated Public Schools	0	0	3.23	6.45	22.58	67.42

Table 6: Distribution of Students Lost due to Vouchers, 1999-2002

(More Treated, Somewhat Treated and Less Treated Schools)

	Loss of Voucher Students (Normalization: Less Treated=1.00)			
	1999	2000	2001	2002
More Treated	1.43	2.09	1.71	1.51
Somewhat Treated	1.37	1.45***	1.35*	1.27
Less Treated	1.00†	1.00†††	1.00†††	1.00†††

*, **, ***: more treated significantly different from somewhat treated at the 10, 5, and 1 percent level, respectively. †,††,†††: more treated significantly different from less treated at 10, 5 and 1 percent level respectively.

Table 7: Effect of Voucher Program on Treatment Status, Milwaukee Phases I and II

Panel A: Phase I	WRCT (% Above)			WRCT (% Below)		ITBS Reading		ITBS Math		ITBS Language Arts	
	OLS	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treated * program	2.44	2.67		-2.10		-2.70		-0.95		-10.84	
	(2.39)	(2.22)		(1.65)		(4.66)		(2.66)		(5.06)	
Treated * program * trend	0.23	0.37		0.25		1.18		0.70		6.55	
	(0.44)	(0.39)		(0.28)		(1.88)		(0.51)		(3.99)	
Treated * program * 1 year			1.37		0.02		3.14		-2.45		-4.46
			(2.31)		(1.82)		(4.57)		(2.73)		(2.65)
Treated * program * 2 years			3.85*		-3.86**		7.26		5.48		2.02
			(2.08)		(1.67)		(5.29)		(3.28)		(3.57)
Treated * program * 3 years			3.07		-1.49		4.96		0.73		
			(2.55)		(2.24)		(5.35)		(3.09)		
Treated * program * 4 years			7.25***		-3.95**				0.41		
			(2.25)		(1.79)				(2.52)		
Observations	1195	1195	1195	1195	1195	717	717	1127	1127	409	409
R-squared	0.16	0.49	0.56	0.46	0.52	0.55	0.54	0.58	0.59	0.69	0.69
p-value	0.12	0.04	0.02	0.38	0.03	0.43	0.57	0.24	0.34	0.10	0.10

Panel B: Phase II	WKCE Reading			WKCE Language Arts		WKCE Math		WKCE Science		WKCE Social Studies	
	OLS	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treated * program	2.98	2.58		4.22**		4.51*		5.65***		4.89***	
	(2.70)	(1.98)		(1.97)		(2.42)		(2.04)		(1.88)	
Treated * program * trend	0.21	0.43		0.40		0.47		1.20*		0.59	
	(0.82)	(0.60)		(0.63)		(2.39)		(1.68)		(0.64)	
Treated * program * 1 year			2.71		4.89***		6.50***		7.90***		5.54***
			(1.71)		(1.60)		(2.04)		(1.64)		(1.56)
Treated * program * 2 years			5.10**		6.66***		6.54***		8.83***		7.03***
			(1.82)		(1.90)		(2.12)		(1.73)		(1.54)
Treated * program * 3 years			4.18**		5.70***		9.35***		8.81***		6.78***
			(1.77)		(1.96)		(2.10)		(1.91)		(1.87)
Treated * program * 4 years			4.79		4.69***		8.24***		10.78***		7.18***
			(1.72)		(1.73)		(2.10)		(1.91)		(1.90)
Observations	669	669	669	669	669	670	670	670	670	670	670
R-squared	0.41	0.79	0.79	0.76	0.77	0.80	0.81	0.82	0.83	0.80	0.81
p-value	0.09	0.01	0.04	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. [1] p-value of the F-test of joint significance of treated shift coefficicents. Huber-White standard errors are in parenthesis. Columns (1), (2), (4), (6), (8), (10) include a time trend, program dummy, program dummy interacted with trend, while columns (3), (5), (7), (9), (11) include year dummies. All regressions include race, sex, percentage of students eligible for free and reduced price lunches and real per pupil expenditure and are weighted by the number of students tested. The FE columns include school-fixed effects; the OLS columns include the treated dummy.

Table 8: Effect of Voucher Program on Treatment Status, Milwaukee Phase I

	WRCT				ITBS					
	WRCT (% above)		(% below)		Reading		Math		Lang. Arts	
	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Somewhat treated * program	3.50		-3.72*		3.21		0.39		-7.95	
	(2.59)		(1.94)		(5.45)		(2.81)		(5.40)	
More treated * program	2.85		-1.60		3.40		-2.97		-12.69***	
	(3.32)		(2.56)		(5.79)		(3.13)		(6.33)	
Somewhat treated * prog * trend	0.64		-0.26		1.22		0.61		6.28	
	(0.47)		(0.34)		(2.02)		(0.54)		(3.62)	
More treated * program * trend	0.67		0.14		3.40		0.75		6.79	
	(0.62)		(0.46)		(5.79)		(0.63)		(4.30)	
Somewhat treated * 1 year after		2.03		-0.54		4.15		-1.35		-0.88
		(2.81)		(2.05)		(4.49)		(2.94)		(2.82)
Somewhat treated * 2 years after		5.38**		-4.45*		7.83		6.14*		5.03
		(2.43)		(1.88)		(5.17)		(3.38)		(3.64)
Somewhat treated * 3 years after		5.01		-2.60		6.78		2.47		
		(3.03)		(2.30)		(5.31)		(3.31)		
Somewhat treated * 4 years after		9.62***		-4.79***				2.62		
		(2.65)		(1.79)				(2.64)		
More treated * 1 year after		-0.92		1.55		1.12		-4.02		-7.86**
		(3.33)		(2.50)		(3.86)		(3.26)		(3.24)
More treated * 2 years after		6.06*		-4.16***		6.59		4.36		0.06
		(3.14)		(2.50)		(5.15)		(3.83)		(4.12)
More treated * 3 years after		5.69		0.38		2.85		-2.22		
		(3.16)		(3.16)		(5.18)		(3.54)		
More treated * 4 years after		11.02***		-4.64*				-3.62		
		(3.34)		(2.60)				(3.13)		
Observations	1195	1195	1195	1195	717	717	1127	1127	409	409
R-squared	0.50	0.58	0.47	0.52	0.55	0.55	0.58	0.60	0.70	0.70
p-value[1]	0.06	0.02	0.82	0.07	0.68	0.62	0.49	0.28	0.04	0.04

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. [1] p-value of the F-test of joint significance of more treated shift coefficicents. Huber-White standard errors are in parenthesis. This table uses the 66-47 sample. Odd numbered columns include a time trend, program dummy, program dummy interacted with trend, while even numbered columns include year dummies. All regressions include school fixed effects, race, sex, percentage of students eligible for free and reduced price lunches and real per pupil expenditure and are weighted by the number of students tested.

Table 9: Effect of Voucher Program on Treatment Status, Milwaukee Phase II

	Reading		Lang. Arts		Math		Science		Soc. Stud.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Less treated * program	2.26		2.92		2.13		1.58		
	(3.32)		(3.32)		(3.21)		(3.14)		
Somewhat treated * program	2.66		5.10**		5.73		7.42***‡		
	(2.15)		(2.14)		(2.37)		(2.30)		
More treated * program	2.35		3.08		3.20		4.63		
	(2.65)		(2.65)		(3.09)		(2.84)		
Less treated * program * trend	-0.46		-0.19		0.44		1.24		
	(1.07)		(1.13)		(1.05)		(1.09)		
Somewhat treated * prog * trend	0.34		0.13		1.17		0.92		
	(0.66)		(0.69)		(0.76)		(0.79)		
More treated * program * trend	1.33†		1.40‡		-2.19		1.76		
	(0.76)		(0.82)		(2.52)		(0.91)		
Less treated * 1 year		2.93		4.69*		5.18*		5.28**	4.39
		(2.83)		(2.63)		(2.71)		(2.54)	(2.54)
Less treated * 2 years		-0.15		1.80		1.40		2.95	4.13*
		(2.52)		(2.41)		(2.70)		(2.43)	(2.24)
Less treated * 3 years		1.26		1.59		4.24*		4.68	4.31***
		(2.25)		(2.33)		(2.56)		(2.83)	(2.52)
Less treated * 4 years		0.53		2.32		5.07*		7.78***	5.93*
		(2.93)		(2.98)		(2.94)		(3.04)	(3.09)
Somewhat treated * 1 year		2.66		5.28***		8.04***		9.07***	5.95
		(1.86)		(1.75)		(2.01)		(1.87)	(1.68)
Somewhat treated * 2 years		4.36***‡‡		7.19***‡‡		8.36***‡‡‡		10.25***‡‡‡	7.66***
		(1.99)		(2.12)		(2.19)		(1.88)	(1.69)
Somewhat treated * 3 years		3.66*		5.30***‡		9.99***‡‡		10.18**‡‡	7.45***
		(1.89)		(2.02)		(2.03)		(2.11)	(1.94)
Somewhat treated * 4 years		3.55*		4.44*		9.35***		11.02***	7.17***
		(1.94)		(1.93)		(2.16)		(2.25)	(2.12)
More treated * 1 year		2.67		4.30*		4.08		7.27***	3.10
		(2.37)		(2.27)		(2.89)		(2.39)	(2.61)
More treated * 2 years		6.50**††		8.37***††		5.75**		9.46***††	5.21**
		(2.41)		(2.61)		(2.86)		(2.52)	(2.56)
More treated * 3 years		6.89**††		8.84***††		8.62***†		8.96**	5.04*
		(2.55)		(2.94)		(3.02)		(2.81)	(3.04)
More treated * 4 years		6.48**††		6.89***		7.88***		12.16***	5.50*
		(2.20)		(2.32)		(2.86)		(2.66)	(2.83)
Observations	669	669	669	669	670	670	670	670	670
R-squared	0.79	0.80	0.77	0.77	0.80	0.82	0.82	0.83	0.82
p-value[1]	0.00	0.01	0.00	0.00	0.35	0.02	0.00	0.00	0.22

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. †,††,†††: more treated significantly different from less treated at 10, 5 and 1 percent level respectively. +,++,+++: more treated significantly different from somewhat treated at 10, 5 and 1 percent level respectively. ‡,‡‡,‡‡‡: somewhat treated significantly different from less treated at 10, 5 and 1 percent level respectively. [1] p-value of the F-test of joint significance of more treated shift coefficicents. Huber-White standard errors are in parenthesis. This table uses the 66-47 sample. All regressions include school fixed effects and are weighted by the number of students tested and control for race, sex, percentage of students eligible for free or reduced price lunches and real per pupil expenditure. Odd numbered columns include time trend, program dummy, interaction of program dummy with trend. Even numbered columns include year dummies. Column (5) includes interactions of trend with treated dummies and columns (6) and (9) include interaction of D_1 dummy ($D_1 = 1$ if year > 1997) with treated dummies.

Table 10: Effect of Voucher Program in Phase I Using a Continuous Treatment Variable
(Sample of Milwaukee Public Schools)

	WRCT				ITBS					
	% Above		% Below		Reading		Math		Language Arts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment * program	0.09		-0.04		0.07		-0.09		-0.20*	
	(0.08)		(0.06)		(0.09)		(0.07)		(0.11)	
Treatment * program * trend	0.00		0.00		-0.02		0.00		0.00	
	(0.02)		(0.01)		(0.04)		(0.01)		(0.06)	
Treatment * 1 year after		-0.09		0.03		0.03		-0.11		-0.17**
		(0.09)		(0.06)		(0.07)		(0.07)		(0.07)
Treatment * 2 years after		0.07		-0.05		0.06		0.00		-0.10
		(0.09)		(0.06)		(0.09)		(0.08)		(0.09)
Treatment * 3 years after		0.03		0.04		-0.02		-0.13		
		(0.11)		(0.08)		(0.09)		(0.08)		
Treatment * 4 years after		0.14		-0.08				-0.15		
		(0.10)		(0.06)				(0.07)		
Observations	920	920	920	920	708	708	1119	1220	441	443
R-squared	0.47	0.55	0.44	0.50	0.53	0.54	0.58	0.54	0.68	0.67
p-value[1]	0.28	0.13	0.75	0.32	0.74	0.76	0.25	0.13	0.01	0.05

Table 11: Effect of Voucher Program in Phase II Using a Continuous Treatment Variable
(Sample of Milwaukee Public Schools)

	Dependent Variable: WKCE Scores									
	Reading		Language Arts		Math		Science		Social Studies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment * program	0.01		0.01		0.01		0.06		-0.02	
	(0.07)		(0.07)		(0.08)		(0.08)		(0.07)	
Treatment * program * trend	0.04*		0.03		-0.09		0.01		-0.13	
	(0.02)		(0.02)		(0.08)		(0.02)		(0.07)	
Treatment * 1 year after		0.01		0.02		-0.03		0.06		-0.02
		(0.06)		(0.06)		(0.08)		(0.06)		(0.07)
Treatment * 2 years after		0.14**		0.15**		0.07		0.12**		0.00
		(0.06)		(0.06)		(0.08)		(0.06)		(0.07)
Treatment * 3 years after		0.14***		0.18***		0.10		0.10		0.02
		(0.06)		(0.06)		(0.08)		(0.07)		(0.08)
Treatment * 4 years after		0.14***		0.12**		0.05		0.11		-0.01
		(0.06)		(0.06)		(0.08)		(0.07)		(0.08)
Observations	509	509	509	509	510	510	510	510	510	510
R-squared	0.75	0.75	0.73	0.74	0.74	0.77	0.77	0.80	0.77	0.78
p-value[1]	0.01	0.02	0.05	0.01	0.44	0.34	0.26	0.31	0.16	0.99

Notes for tables 10 and 11: *, **, ***: significant at the 10, 5, and 1 percent level, respectively. [1] p-value of the F-test of joint significance of shift coefficients due to treatment. Treatment instensity is proxied by the percentage of students eligible for free or reduced price lunches. Huber-White standard errors are in parentheses. All regressions include school fixed effects and are weighted by the number of students tested and control for race, sex and percentage of students eligible for free or reduced-price lunches. Odd numbered columns include time trend, program dummy, interaction of program dummy with trend. Even numbered columns include year dummies. In table 11, columns (5) and (9) include interactions of trend with treatment ($\%frl$) and columns (6) and (10) include interaction of D_1 dummy ($D_1 = 1$ if year > 1997) with treatment.

Table 12: Comparing the Impact of the Milwaukee Voucher Program in Phase I with that in Phase II

	Reading				Language Arts		Math	
	Phase I			Phase II	Phase I	Phase II	Phase I	Phase II
	WRCT		ITBS	WKCE	ITBS	WKCE	ITBS	WKCE
	% Above	% Below						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A	Using All Treated Schools as a Group							
Treated * 1 year	0.09	0.00	0.17	0.21	-0.27*	0.38***	-0.15	0.43***
Treated * 2 years	0.24*	-0.32**	0.39	0.39**	0.12	0.52***	0.33	0.44***
Treated * 3 years	0.19	-0.13	0.27	0.32**		0.44***	0.04	0.62***
Treated * 4 years	0.45**	-0.36**		0.37***		0.36***	0.02	0.55***
Panel B	Using Treatment Groups							
More Treated * 1 year after prog	-0.06	0.14	0.06	0.20	-0.48**	0.33**	-0.24	0.27
More treated * 2 years after prog	0.38*	-0.38*	0.36	0.50**	0.00	0.65**	0.26	0.38**
More treated * 3 years after prog	0.35	0.03	0.15	0.53**		0.69***	-0.13	0.57***
More treated * 4 years after prog	0.69*	-0.42*		0.50**		0.53***	-0.22	0.52***
Panel C	Using Treatment Groups, After Correcting for Mean Reversion							
More Treated * 1 year after prog	-0.06	0.14	0.06	0.20	-0.48**	0.33**	-0.24	0.03
More treated * 2 years after prog	0.38*	-0.38*	0.36	0.50**	0.00	0.65**	0.26	0.14**
More treated * 3 years after prog	0.35	0.03	0.15	0.53**		0.69***	-0.13	0.33***
More treated * 4 years after prog	0.69*	-0.42*		0.50**		0.53***	-0.22	0.28***
Panel D	Using Continuous Treatment Variable							
More Treated * 1 year after prog	-0.005	0.003	0.002	0.001	-0.01**	0.002	-0.006	0.002
More treated * 2 years after prog	0.004	0.004	0.003	0.01**	-0.006	0.01**	0.00	0.005
More treated * 3 years after prog	0.002	0.003	-0.001	0.01***		0.01***	-0.008	0.007
More treated * 4 years after prog	0.009	0.007		0.01***		0.01***	-0.009**	0.003

All figures are in terms of respective sample standard deviations and pertain to the 66-47 sample. All figures are obtained from regressions that contain school fixed effects, year dummies, interactions of year dummies with more treated, somewhat treated, less treated year dummies respectively, are weighted by the number of students tested and control for race, sex, percentage of students eligible for free or reduced price lunches and real per pupil expenditure. For Panels A, B and C samples: Standard deviation of WRCT % above (% below) scores = 16 (10.98), Standard deviation of ITBS Reading scores = 18.45, Standard deviation of ITBS Language Arts scores = 16.23, Standard deviation of ITBS Math scores = 16.71, Standard deviation of WKCE Reading scores = 13.07, Standard deviation of WKCE Language Arts scores = 12.88, Standard deviation of WKCE Math scores = 15.01. For Panel D sample: Standard deviation of WRCT % above (% below) scores = 15.81 (11.56), Standard deviation of ITBS Reading scores = 18.45, Standard deviation of ITBS Language Arts scores = 16.23, Standard deviation of ITBS Math scores = 16.71, Standard deviation of WKCE Reading scores = 12.92, Standard deviation of WKCE Language Arts scores = 13.08, Standard deviation of WKCE Math scores = 14.44.

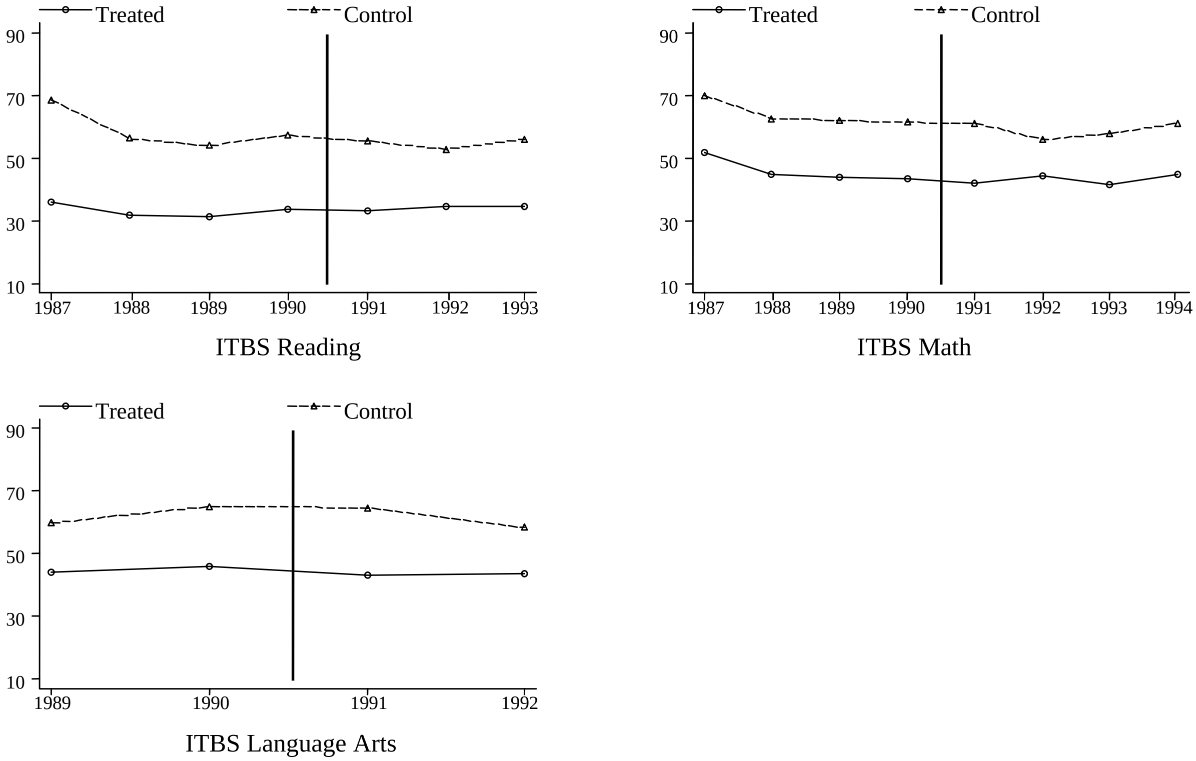

Figure 1. Panel A. Milwaukee Voucher Program, Phase I
(Means)

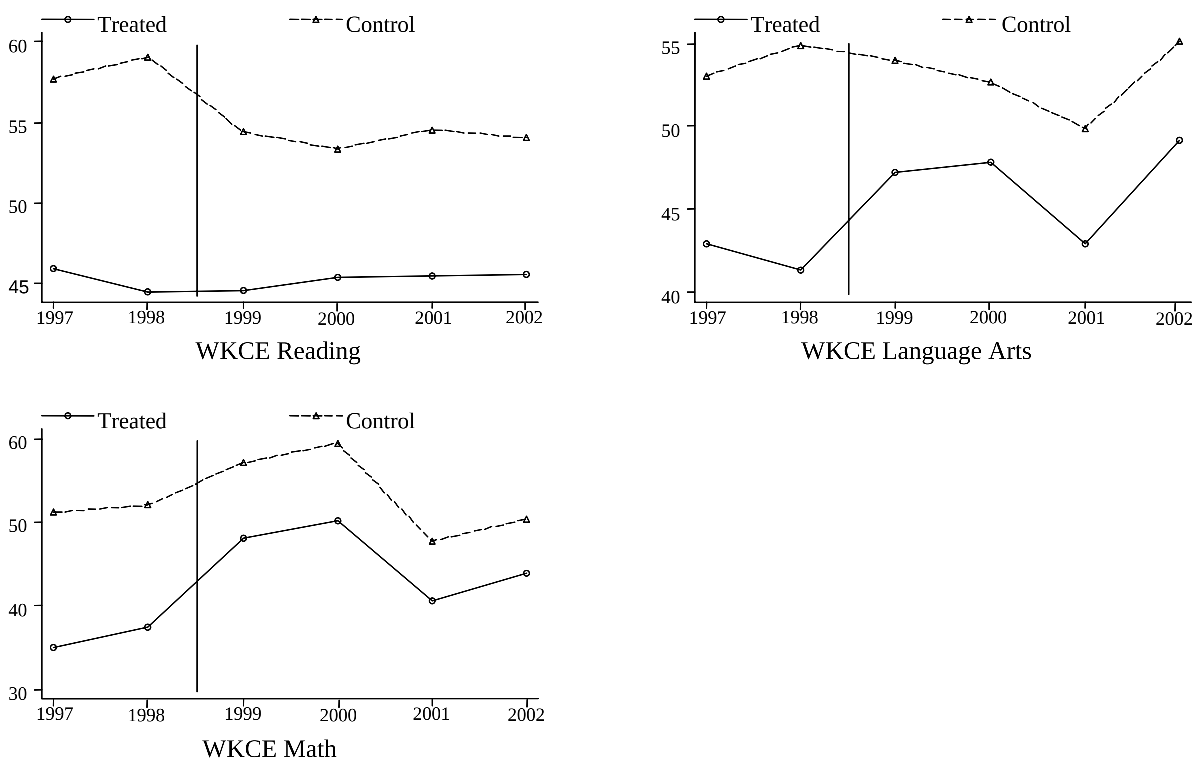

Figure 1. Panel B. Milwaukee Voucher Program, Phase II
(Means)

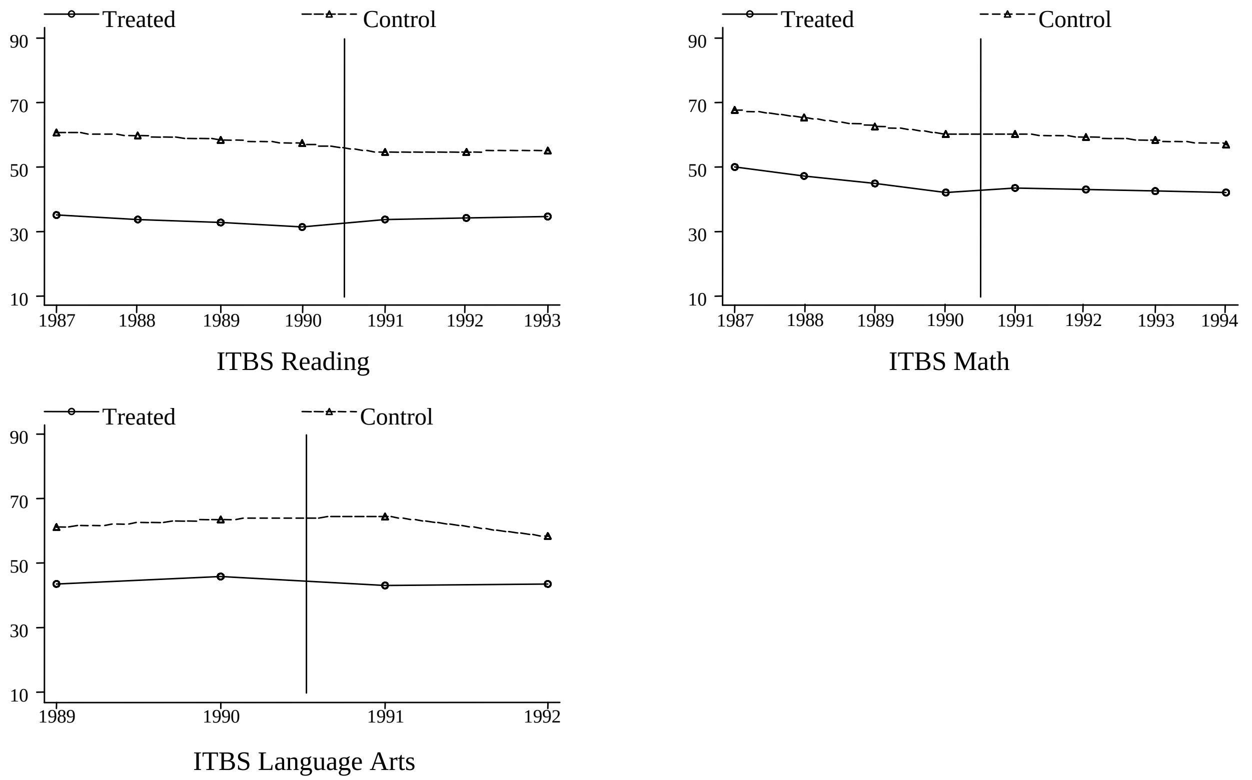

Figure 2. Panel A. Milwaukee Voucher Program, Phase I

(Predicted values from OLS regressions corresponding to Model 1, sample of treated and control schools)

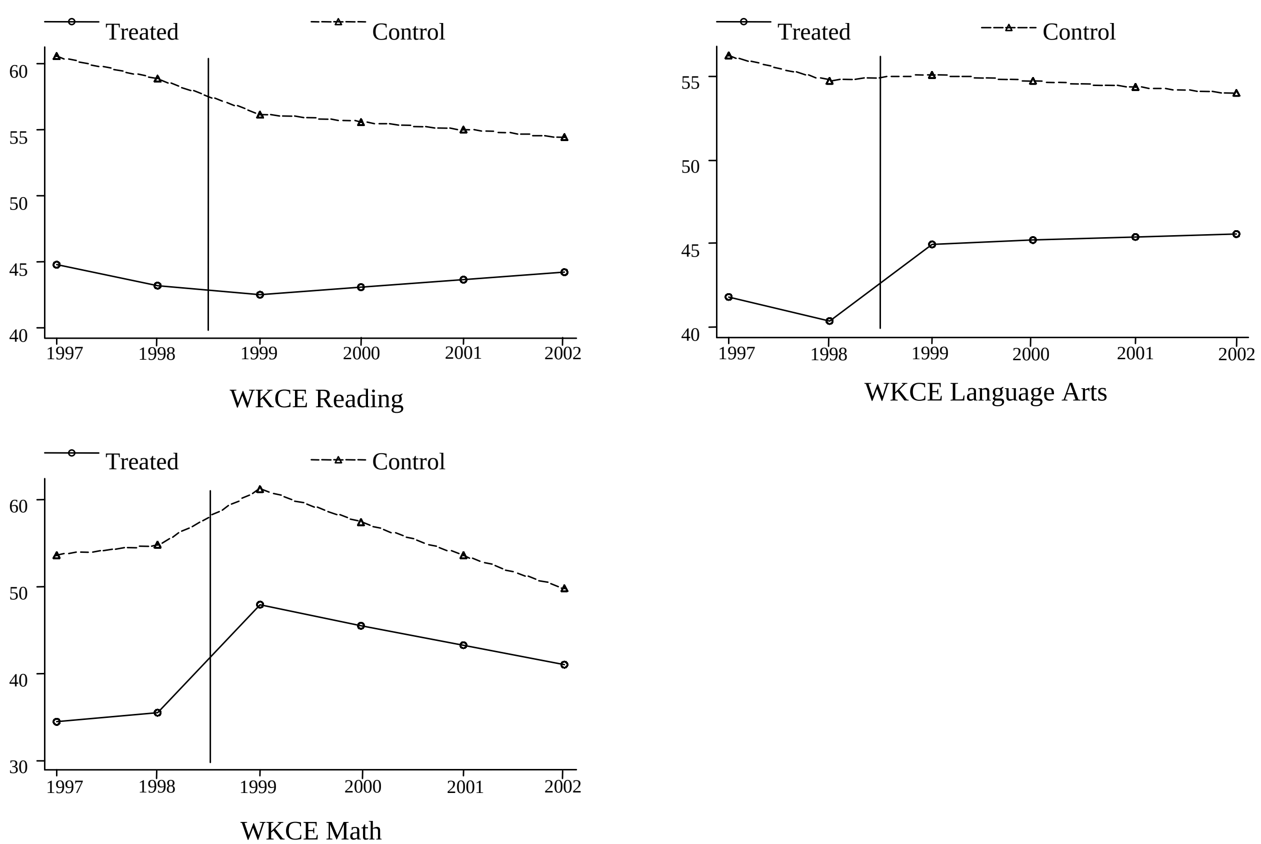

Figure 2. Panel B. Milwaukee Voucher Program, Phase II

(Predicted values from OLS regressions corresponding to Model 1, sample of treated and control schools)

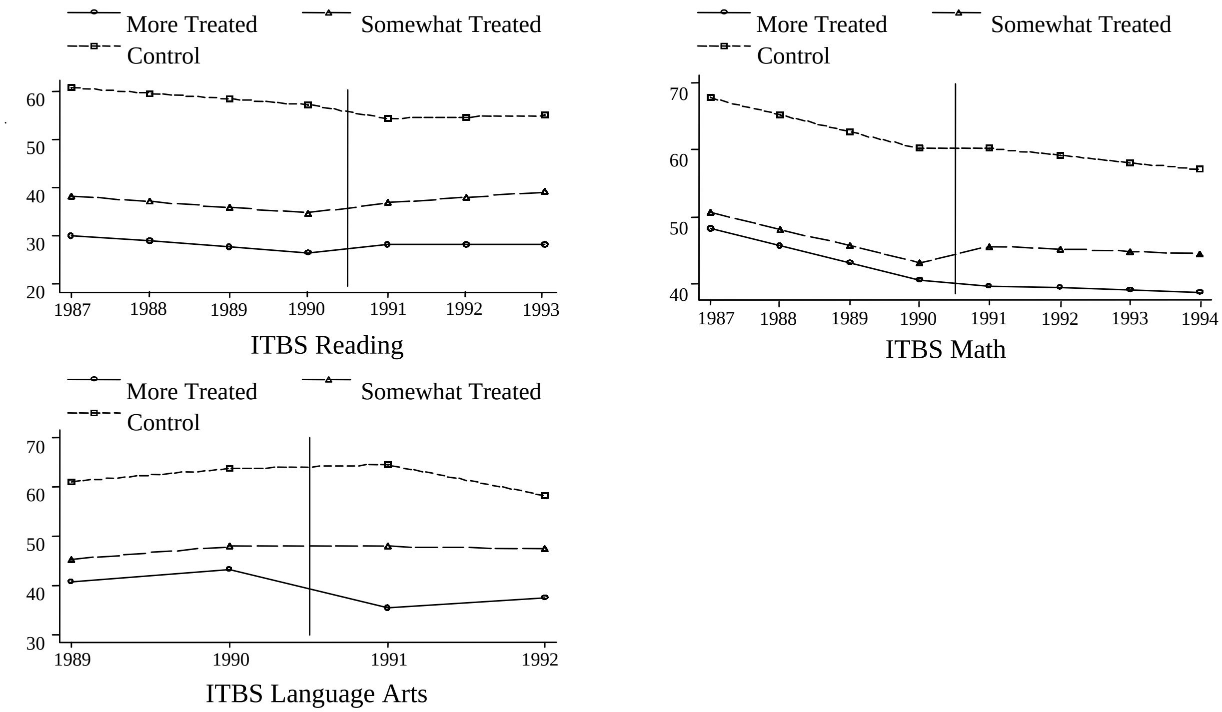

Figure 3. Panel A. Milwaukee Voucher Program, Phase I

(Predicted values from OLS regressions corresponding to Model 1, sample of more treated, somewhat treated, and control schools)

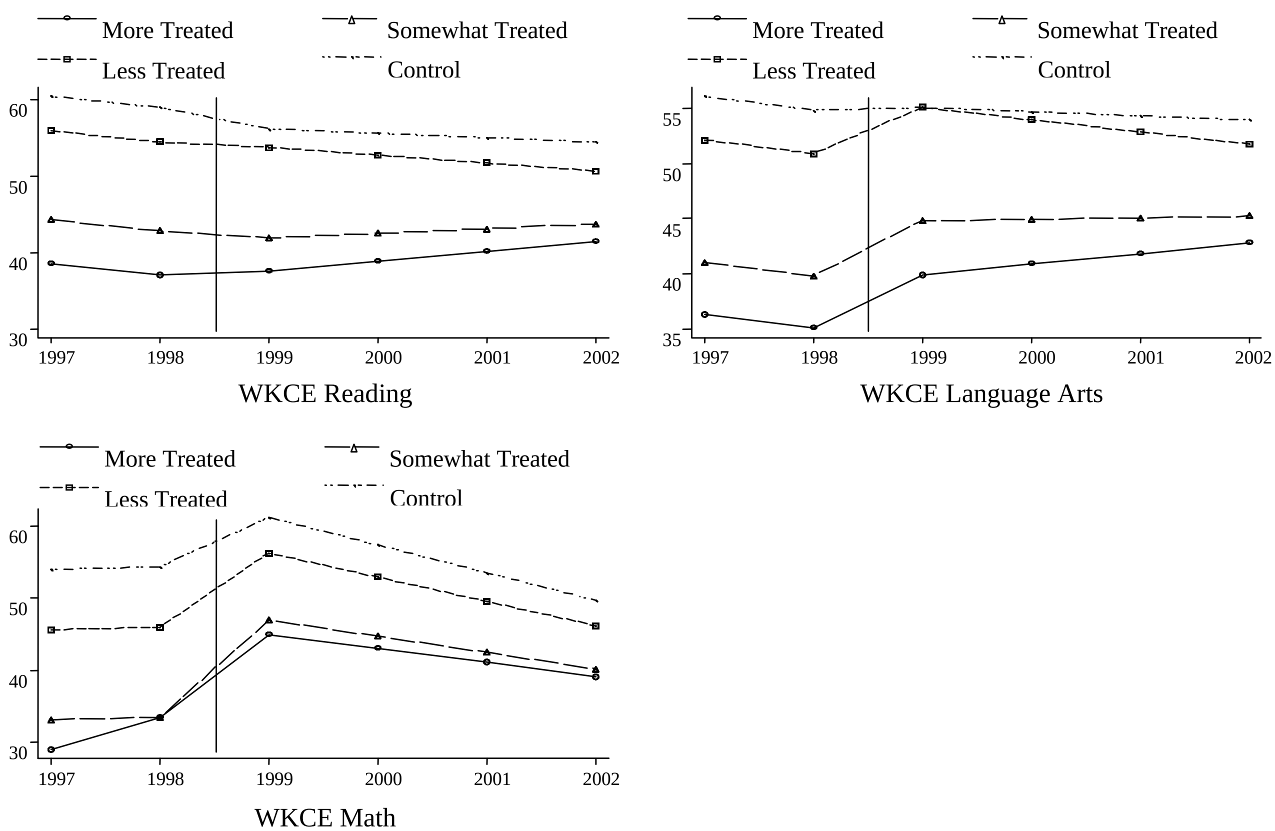

Figure 3. Panel B. Milwaukee Voucher Program, Phase II

(Predicted values from OLS regressions corresponding to Model 1, sample of more treated, somewhat treated, less treated, control)