Federal Reserve Bank of New York
Staff Reports

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

Rajashri Chakrabarti

Staff Report no. 300
September 2007

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 Staff Reports, no. 300
September 2007
JEL classification: H4, I21, I28

Abstract

The Milwaukee voucher program, as implemented in 1990, allowed only nonsectarian private schools to participate in the program. However, following a Wisconsin Supreme Court ruling, the program was expanded to include religious private schools in 1998. This second phase of the voucher program 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, because of some changes in funding provisions, the revenue loss per student from vouchers increased in the second phase of the program. This paper analyzes, both theoretically and empirically, the effects of these changes on public school performance (as measured by test scores) in Milwaukee. It argues that voucher design matters and that the choice of parameters in a voucher program is crucial in determining the effects of public school incentives and performance. In the context of a theoretical model of public school and household behavior, the paper establishes that the policy changes will lead to an improvement of the public schools in the second phase of the program as compared with the first phase. Following Hoxby (2003a, 2003b) in the classification of treatment and control groups and using 1987-2002 data and a difference-in-differences estimation strategy in trends, the paper then shows that the theoretical prediction is validated empirically. This result is robust to alternative samples and specifications and survives robustness checks, including correcting for mean reversion.

Key words: vouchers, public school performance, competition, mean reversion

---

Chakrabarti: Federal Reserve Bank of New York (e-mail: rajashri.chakrabarti@ny.frb.org). The author is grateful to Steve Coate, Ron Ehrenberg and Miguel Urquiola for helpful comments, suggestions and encouragement. She thanks Robin Boadway, Julie Cullen, Caroline Hoxby, Brian Jacob, George Jakubson, Paul Peterson, and seminar participants at Harvard University, the Econometric Society Winter Meetings, and the 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 its 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.

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 differential effects of alternative voucher programs is key to designing an effective voucher policy. This paper contributes in this direction by studying, both theoretically and empirically, how changes in some crucial policy parameters midway through the implementation of the Milwaukee voucher program affected public school incentives and performance (as measured by test scores). It provides strong evidence that voucher design matters as far as public school response is concerned.

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 schools participated for the first time in the school year 1998-99. This led to a massive 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 starting from the school year 1999-2000.

This paper analyzes the effects of these shifts on the incentives and responses of the treated public schools in Milwaukee. 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 Milwaukee program.

The paper develops its argument in the context of a formal theoretical model that is designed to capture the basic features of the Milwaukee voucher program. It has three agents: – the public school, the households and the private schools. The demand for public school is endogenously determined from equilibrium household behavior, giving micro-foundations to the public school payoff function. In an equilibrium framework of public school and household behavior, the model endogenously determines public school quality and its ingredients – public school effort and peer group quality. It yields an empirically testable prediction—the treated public schools will exhibit unambiguously higher improvement in the second phase of the program as compared to the first phase.

Using school-level test score data from Wisconsin, the paper next proceeds to test the theoretical prediction. Implementing a difference-in-differences estimation strategy in trends, it estimates the program effects in each of the first and second phases of the program by comparing the improvement of the treated schools in each phase with an appropriate set of control schools. Controlling for potentially confounding pre-program time trends and post-program common shocks, the paper finds considerable evidence in favor of the theoretical prediction. This finding is robust to alternative strategies and specifications and continue to hold after controlling for other confounding factors such as mean reversion. Moreover, it has strong policy implications.

The study of school vouchers has attracted considerable attention in the last decade. Nechyba (1996, 1999, 2000) analyzes distributional effects of alternative voucher policies in a general equilibrium framework that endogenizes residential choice. Epple and Romano (1998) argue that vouchers lead to sorting by income and ability. They model private school and household behavior, but assume public schools to be passive. Epple and Romano (2002) examine how alternative voucher designs can affect stratification and technical efficiency. They allow for public school technical inefficiencies, but these inefficiencies are taken to be exogenous in their study. In particular, none of the above studies endogenize public school quality.

Nechyba (2003) allows for efficiency gain in the public schools facing competition from vouchers.[1] However, he does not model public school behavior. Manski (1992) considers the impact of vouchers

[1] He includes two constants in the public school production function that exogenously increase with a decrease in peer quality variance and an increase in the share of private school attendance respectively.

on public school expenditure and social mobility, while allowing for rent-seeking public schools. But unlike the present paper, understanding the effect of changes of different policy parameters in a voucher program on public school performance is not a concern in Manski. Modeling public school quality, McMillan (2004) shows that under certain circumstances, public schools may find it optimal to reduce productivity when a voucher is introduced. The first difference once again is that the focus of this paper is to analyze the impact of changes in alternative policy parameters in a voucher program on public school performance,—this is not a concern in McMillan. This paper shows that even within a traditional voucher program (the type considered by both Manski and McMillan), different choices of parameters can have radically different effects on public school response. Second, while McMillan employs a theoretical approach, this paper first tries to understand the effect of the changes in policy parameters in a theoretical context and then seeks to test the intuitions obtained there in an empirical framework. Third, unlike McMillan, this paper models peer quality which is considered to be an essential component of school quality. Fourth, it derives the demand for public school from equilibrium household behavior, thus providing microfoundations to the public school payoff function, unlike McMillan.

A number of empirical studies look at the effect of vouchers on the performance of students who move to private schools with vouchers (the "choice students"). For a comprehensive review of this literature, see Hoxby (2003b) and Rouse (1998). The empirical literature on the impact of vouchers on public school performance in the U.S. has been relatively sparse. Greene (2001, 2003) finds positive effect of the Florida program on the performance of treated schools. However, the classification into different treatment groups in Greene (2003) is based on post-program grades of schools and hence is susceptible to the endogeneity problem. Moreover, Greene (2003) uses a difference-in-differences analysis and takes all other schools (other than the treatment group) as the control group. Since this comparison group differs considerably from the treatment groups in terms of demographic characteristics, school scores and grades, it is not clear how appropriate a control group it is. In response to Greene's (2001) paper, a spurt of studies took place (Camilli and Bulkley (2001), Harris (2001), Kupermintz (2001)) that express doubt that the program effect in the Greene study is contaminated by mean reversion[2] and/or stigma

[2] For a discussion of the mean reversion problem, see Chay et al. (2003).

effect of getting the lowest performing grade "F". However, it is not clear that the above studies in response to Greene were able to purge the program effect of mean reversion.[3] Moreover, unlike the present study, none of them control for any pre-program differences in trend between treatment and control schools which may bias the program effects.

By far the most substantial contribution to this still sparse literature is to be found in the important studies by Hoxby (2003a, 2003b) on the Milwaukee voucher program. This study has been greatly informed by the Hoxby studies. Hoxby (2003a) analyzes the impact of the Milwaukee voucher program on public schools after the Wisconsin Supreme Court ruling of 1998. She uses a novel strategy to pick her treatment groups of schools. Since the MPS students eligible for free or reduced price lunches were the ones eligible for vouchers (see footnote 30), 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 are potentially affected by the program, she chooses, as her control group, a set of schools within Wisconsin but outside Milwaukee that are most similar to the Milwaukee schools. (Her treatment-control strategy is discussed in more detail in section 6.1.1.) 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[4] (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.

[3] In his study, Greene (2001) argues that mean reversion is not a problem as the gains achieved by low scoring F schools are similar to those of the high scoring F schools between 1999 and 2000. However, similar gains of low scoring and high scoring F schools do not imply an absence of mean reversion since 2000 is a post-program year. In fact, even in the presence of mean reversion, the coefficients of the high scoring and low scoring F schools can be similar if there are differential program effects between these two groups. The studies in response to Greene (2001) seek to arrive at mean reversion corrected program effect by subtracting the post-program (2000) score from the predicted score in 2000, where the predicted score is obtained from a regression of the 2000 score on the pre-program (1999) score. However, in this strategy, the mean reversion effect is confounded with the program effect (since 2000 is a post-program year) and the mean reversion correction gets rid of at least part of the program effect. (Harris (2001) and Kupermintz (2001) exclude the F schools in their predicted score regressions. However, it is not clear that any mean reversion effect from the other groups of schools can be attributed to the F schools.)

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

This paper follows Hoxby in the treatment-control group classification. As in Hoxby, it uses the percentage of free or reduced price lunch eligible students in the Milwaukee public schools to classify them into different 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, 2003b) 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,—more specifically, in investigating whether a voucher program that is 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, this study employs two alternative methods of sample formation. Following Hoxby, this study classifies Milwaukee schools into different treatment groups based on the percentages of their students eligible for free or reduced price lunches. However, this study classifies the Milwaukee schools into three treatment groups (unlike two in Hoxby) so that, on the one hand, the treatment groups are more homogenous while on the other they are starker from each other. Moreover, to test the robustness of the results, it also considers different samples that are constructed by varying the cutoffs that divide the Milwaukee schools into different treatment groups. I follow Hoxby in the control group classification also, although there are differences as outlined in section 6.1.1. One disadvantage of the above treatment group strategy is that it constrains the program effect to be the same for all schools within a treatment group, that is, it does not allow within group differences in treatment intensities to affect performance. Therefore I also consider an alternative strategy. This second strategy uses a continuous treatment variable where the intensity of treatment is proxied by the schools' percentage of free or reduced price lunch eligible students in 1990.[5]

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 both before the program (1990) and before the program shift (1998), a potential concern is that any improvement of the schools may

[5] However, a disadvantage of this strategy is that it assumes that the program effect varies linearly with treatment intensity. To address this problem, I also estimate alternative specifications that alow for nonlinearities in the above relationship. These are discussed in detail in section 6.1.

be due to regression to the mean rather than a program effect. Fourth, 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, after the program shift as well as after the initial program. This is instructive since the public school response may vary across the different years after the program. Fifth, unlike Hoxby, this study controls for the possibility that changes in student composition of schools may bias the estimated effects of voucher competition.

Sixth, unlike Hoxby, this study also investigates whether the potential competition faced by the Milwaukee schools was actually effective. The Milwaukee schools, especially the more treated ones, had a substantial proportion of their students eligible for vouchers. But this competition would not be functional if there were not enough private schools surrounding them. To investigate the extent of effective competition, I investigate the distribution of choice schools around Milwaukee public schools and also the extent of their actual loss of students to the program. Finally, an important difference with Hoxby as well as the other studies mentioned above is that in addition to an empirical part, this paper contains a theoretical part that is designed to capture the basic features of the Milwaukee program. It aims to get an intuition into the nature of incentives created by the program shifts and seeks to analyze their effects on public school performance. The empirical part seeks to test the theoretical predictions.

Although there are multiple papers that analyze the effects of alternative voucher policies on stratification, distribution and welfare[6], 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

[6] Nechyba (2000) and Caucutt (2002) examine distributional and welfare consequences of targeting vouchers to low income types; Epple and Romano (2002) and Hoxby (2001) consider the effect of alternative voucher policies on stratification and equity. These papers relate to voucher design, but their concern is not its impact on public school performance.

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.

2 Institutional Background—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 and student participation in the MPCP was limited to 1% of the MPS membership.[7] The 1993 Wisconsin Act 16 raised the 1% limit to 1.5% which was further raised to 15% in 1996-97 under the 1995 Act 27.

The 1995 Act 27 also proposed that private sectarian or religious schools be allowed to participate in the program. But this change was immediately challenged in court and a preliminary injunction prohibiting this change was issued. On June 10, 1998, in a landmark decision (Jackson v. Benson), the Wisconsin Supreme court ruled 4-2 that including religious schools in choice is constitutional. The decision became even more important when on November 9, 1998, the U.S. Supreme court declined (by an 8-to-1 vote), without comment, to review the case, ensuring that the ruling would stand in the foreseeable future. Following the Wisconsin Supreme Court ruling, the religious schools participated in the program for the first time in the 1998-99 school year.

In spite of the cap on MPCP enrollment, as mentioned above, this participation constraint was not binding. As table 1 shows, the number of applicants was almost always considerably less than that allowed by the program during the years 1990-91 through 1994-95. The picture has been very similar after 1994-95. 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 were not only well below the number authorized by the statute but were also considerably

[7] Membership is the number of pupils enrolled in the school who are counted for the purpose of computing state general equalization aids. It is based on an average of two counts in the regular school year (September and February) plus a full time equivalency for summer school.

less than the number of applicants in each of the years 1990-91 through 1994-95.

Therefore 1998 constitutes a benchmark year in the history of the MPCP. The Wisconsin Supreme Court ruling of 1998 allowed religious schools to participate in the program thus relaxing 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 the MPS membership fell for the first time in the 1998-99 school year.

The financing of the MPCP has also changed over the years. Under the MPCP, as implemented in the 1990-91 school year, state aid would follow the pupil from the MPS to the private school. Pupils participating in the choice program were then included in the membership count for MPS on a prior year basis even though the pupils 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[8] and the MPCP was funded by reducing the state aid for the MPS district by the voucher amount times the number of students attending under the MPCP.

The 1998 Act 9 made a number of changes with respect to the funding of the program which were implemented in the 1999-2000 school year. The definition of membership was changed to exclude MPCP pupils—unlike earlier, starting from the 1999-2000 school year, 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. Under 2001 Act 16, which was implemented in the 2001-02 school year, the MPCP was funded 45% from a reduction in state aid to MPS and 55% from the state general purpose revenue, so that the other districts did not bear the burden of financing of the MPCP.

[8] More precisely, it was the equalization aid per member.

Although only 50% of the MPCP expenditure (45% from 2001-02) came from the MPS from 1999-2000, 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 per student to the MPS in the period prior to Act 9 was v, while the loss per student was at least $(v + \frac{v}{2})$ from 1999-2000.[9 10]

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 massive 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 1998-99 to 28.7 million in 1999-2000 which was funded entirely by a corresponding reduction in state aid to the MPS. The MPCP amount continued to increase due to the increase in the number of choice students, but the 50% (and latter 45%) 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 from 1999-2000 was much more than illustrated here)—the table only illustrates the distribution of the MPCP burden.[11]

This paper 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 1998-99 following the Wisconsin Supreme Court ruling and the discrete increase in the loss of revenue per student from 1999-2000 when the Act 9 changes in funding were implemented. Note that the 1998-99 year effect would capture the effect of an increased private school participation, while the effect from 1999-2000 onwards would capture the effect of both an increased private school participation and an increase in

[9] This changed to $(v + 0.45v)$ from 2001-02.

[10] 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.

[11] Also note that the voucher amount was larger in the second phase than in the first. This would reinforce the above larger loss in revenue per pupil in the second phase.

per pupil loss in revenue due to vouchers.

3 The Model

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) to all households that choose to attend it. The quality q is a composite of two factors: public school effort and public school peer-group quality. The objective of the public school is to maximize net revenue, which I call "rent" in rest of the paper. I adhere to the general line of thought in the school competition literature [Hoxby (2003a), Manski (1992), McMillan (2004)] that the public school maximizes surplus or net revenue.[12] Rent is simply defined as revenue minus costs. Public school cost is given by: $C_p(N, e) = c_1 + c(N) + C(e)$, where c_1 is a fixed cost, N is the number of students in public school and e is public school effort. Both $c(.)$ and $C(.)$ functions are assumed to be increasing and strictly convex in their respective arguments. Per pupil revenue is denoted by p and is exogenously given. Revenue depends on per pupil revenue, number of students and in the presence of vouchers also on the loss of revenue per student lost due to vouchers. (It is discussed in more detail in section 4.2.) In the absence of vouchers, it is simply given by $p.N$.

There is a continuum of private schools providing a continuum of quality levels. Each private school is "passive" and does not take any maximizing decision.[13] 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. A household obtains utility (U) from the consumption of the numeraire good (x), school quality (θ) and its ability (α). The household utility function is assumed to be continuous and twice differentiable and is given by $U(x, \theta, \alpha) = h(x) + \alpha u(\theta)$. To simplify analysis, I assume $h_{xx} = 0$. The function u is increasing and strictly concave in θ. It follows

[12] An alternative formulation could be to model the public school as a quality maximizer. However, in that case there would be no argument for voucher programs as far as improving public school quality is concerned.

[13] This is in keeping with the feature of the U.S. voucher experiments, by which private schools are not allowed to discriminate between students. They have to accept all students unless oversubscribed and have to accept students randomly when oversubscribed. (Of course, in the voucher experiments, they can choose whether or not to enter. I abstract from that here for simplicity.)

that households with higher ability have a higher preference (marginal valuation) for school quality, $U_{\theta\alpha} > 0$.

School qualities available to a household are public school quality and a continuum of (exogenously given) private school qualities. Public school quality $q = q(e, b)$ is a continuous, twice differentiable, increasing and concave function of public school effort $e \in [e_{min}, e_{max}]$ and public school peer quality b. Public school peer quality is defined as the mean ability of the public school student body.[14] If a public school household decides to switch to a private school with vouchers, it incurs a positive switching or relocation cost c.[15]

The paper models 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 program. The simple public-private system consists of two stages. In the first stage, the public school chooses effort. In stage 2, households choose between schools after observing public school effort.[16] Peer-group quality and public school quality are simultaneously determined.

The Milwaukee program is analyzed in three stages. In the first stage, the government announces voucher v and a cutoff (or target) income level y_T. Only households with $y \leq y_T$ are eligible for vouchers. In stage 2, facing the program, the public school chooses effort. In stage 3, households choose between schools and incur switching costs if they transfer out of public school. Peer-group quality and public school quality are simultaneously obtained.

The number of private school seats available for households applying with vouchers is limited. (It is a binding constraint in the sense that the number of applicants exceed the number of seats.) Private schools pick voucher students randomly so that a random sample of those that apply are selected,—a certain proportion (say, β) of the applicants are accepted/successful and $(1 - \beta)$ proportion are

[14] Public school quality can be thought of as being embodied in public school scores. The notion here is that public school scores reflect both public school effort and public school peer-group quality, which in turn depends on the abilities of the public school students. In other words, both public school characteristics and student characteristics contribute to school scores.

[15] Switching schools may separate children from their friends and social circles, require them to adapt to new teachers or new curriculum, create logistical or scheduling difficulties, interfere with extracurricular activities etc. The switching cost "c" captures these costs.

[16] The general notion in both the systems is that households observe last year's scores and whether vouchers were given and then choose between schools.

rejected/unsuccessful and return to the public school.

Each of the systems constitutes a game between two players: the public school and the households. Consider the Milwaukee-type voucher program. Public school households moving in stage 3 observe the program and public school effort and decide whether or not to apply with vouchers. The public school moving in stage 2 observes the program, correctly anticipates household behavior and makes its rent maximizing effort choice. In stage 1 of the simple public-private system, the public school chooses its rent maximizing effort after correctly anticipating household behavior. In stage 2, households choose between schools after observing public school effort.

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.

4 Characterization of the program equilibria

This section first solves for the household and public school equilibria. Next, it compares the public school qualities under the PP and voucher equilibria. Finally, it analyzes the effect of relaxation of the private school participation constraint and an increase in the revenue loss from vouchers on public school quality in a voucher equilibrium, so as to compare the equilibrium quality under Milwaukee phase I and phase II.

4.1 Household behavior

This subsection analyzes the household behavior under the two systems in a common framework. Each household can either choose a public or a private school. It gets utility $h(y) + \alpha u(q(e, b))$ from a public school and utility $h(y + v - t \cdot Q^* - c) + \alpha u(Q^*)$ from private, where Q^* is the optimal private school quality choice of household (y, α) given v, t and c. The parameter v takes on a value of zero under the pre-program public-private system, and under the Milwaukee voucher system if the income of the household exceeds y_T. On the other hand, v takes on an exogenously given positive value under the voucher program, for all households with $y \leq y_T$. A household (y, α) chooses private school iff $h(y+v-t\cdot Q^*-c)+\alpha u(Q^*) > h(y)+\alpha u(q(e,b))$.[17] Define $D = [h(y+v-t\cdot Q^*-c)+\alpha u(Q^*)]-[h(y)+\alpha u(q(e,b))]$.

Suppose all households expect a peer group quality $b^e \in [0,1]$. [18] Then for each y and given t, v, e, c and expected peer group quality $b^e \in [0,1]$, 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:

$$[h(y+v-t.Q^*-c)+\alpha u(Q^*)]-[h(y)+\alpha u(q(e,b^e))] = 0 \tag{3.1.1}$$

where Q^* is the optimal private school quality choice of the household $(y, \hat{\alpha}(y))$.[19] Since the indirect utility and the q functions are continuously differentiable and $D_\alpha > 0$, $\hat{\alpha} = \hat{\alpha}(y; v, e, b^e, t, c)$ (3.1.1a) can be obtained as a continuously differentiable function of v, b^e, t, c, for each y.[20] Using the implicit function theorem it is straightforward to check that for each income level, the cutoff ability level $\hat{\alpha}$ is decreasing in v and increasing in e, b^e, t and c.

Given b^e, peer group quality b is given by:

$$b = \frac{\int_0^{y_T}\int_0^{\hat{\alpha}(y,b^e,v,.)} \alpha d\alpha dy + (1-\beta)\int_0^{y_T}\int_{\hat{\alpha}(y,b^e,v,.)}^{\hat{\alpha}_0(y)} \alpha d\alpha dy + \int_{y_T}^1\int_0^{\hat{\alpha}(y,b^e,0,.)} \alpha d\alpha dy}{\int_0^{y_T}\int_0^{\hat{\alpha}(y,b^e,v,.)} d\alpha dy + (1-\beta)\int_0^{y_T}\int_{\hat{\alpha}(y,b^e,v,.)}^{\hat{\alpha}_0(y)} d\alpha dy + \int_{y_T}^1\int_0^{\hat{\alpha}(y,b^e,0,.)} d\alpha dy} = g(b^e, e, v, t, c, y_T, \beta) \tag{3.1.2}$$

The numerator denotes the sum of abilities of all households attending public school. The denominator denotes the total number of households attending public school. When $v \neq 0$, the first term in each of numerator and denominator corresponds to low income households who are eligible for vouchers

[17] Note that this holds not only under the simple public-private system but also under the voucher system. Under the voucher system, a household applying with vouchers knows that it will be accepted only with a positive probability (less than one). However, if it prefers private to public, it prefers a lottery between private and public to public and still continues to apply to private school. Formally a voucher eligible household chooses private school iff $\beta[h(y+v-t\cdot Q^*-c)+\alpha u(Q^*)]+(1-\beta)[h(y)+\alpha u(q(e,b))] > [h(y)+\alpha u(q(e,b))] \Rightarrow [h(y+v-t\cdot Q^*-c)+\alpha u(Q^*)] > [h(y)+\alpha u(q(e,b))]$

[18] I assume that there are always some households in the public and some households in the private sector at each income level. This assumption is made for simplicity. All results go through as long as there is at least one income for which this assumption holds.

[19] To save some notation the optimal private school quality choice of the corresponding household is always denoted by Q^*. It is obvious that the value of Q^* will change with income and ability.

[20] Since $h_{xx} = 0$, the income effect of school quality is zero. It follows that, given other parameters, in the voucher system, the cutoff ability level is independent of y in both the ranges $[0, y_T)$ and $(y_T, 1]$, but there is a discontinuity at y_T. In the PP system, the cutoff ability is independent of y in [0,1]. Moreover, although the household equilibrium is characterized by stratification by ability ($\frac{\delta D}{\delta \alpha} > 0$), there is no stratification by income ($\frac{\delta D}{\delta y} = 0$).

but choose public school; the last term corresponds to high income households who choose public school; the middle term corresponds to the proportion of low income households who are unsuccessful in getting a voucher seat (in spite of application) and return to public school. Under the simple public-private system, $v = 0$.[21] At equilibrium, b corroborates the initial conjecture b^e, that is, $b = b^e$. (3.1.3)

The household equilibrium is characterized by (3.1.1)-(3.1.3). If all households expect a peer-group quality, then at equilibrium this expectation has to be fulfilled. Mathematically, given parameters e, v, t, c, y_T, β, a fixed point in b is reached. It can be shown that a household equilibrium always exists.[22] (See appendix for proof.) From (3.1.1)-(3.1.3), the equilibrium peer quality satisfies the equation $b^* = g(b^*, e, v, t, c, y_T, \beta)$. The corresponding equilibrium allocation of households between public and private sectors is characterized by $\hat{\alpha}(y, b^*, .)$ for $y \in [0, 1]$. The number of students in public school at the household equilibrium b^* is given by $N(b^*, e, v, t, c, y_T, \beta) = \int_0^{y_T} \int_0^{\hat{\alpha}(y,b^*,v,.)} d\alpha dy + (1-\beta) \int_0^{y_T} \int_{\hat{\alpha}(y,b^*,v,.)}^{\hat{\alpha}_0(y)} d\alpha dy + \int_{y_T}^1 \int_0^{\hat{\alpha}(y,b^*,0,.)} d\alpha dy$

The equilibrium number of public school students increases with public school effort and decreases with vouchers. (See appendix for proof.) The intuitive argument here is as follows. An increase in public school effort leads to an influx of higher ability households at each income level, that is an increase in the equilibrium cutoff ability level $\hat{\alpha}(y, b^*)$ at each income level. This occurs through two channels. Given peer quality, an increase in e induces households just above the cutoff at each income level to switch to the public school. This increases peer quality, leading to a further influx of higher ability households just above the cutoff from the private to the public sector. These two effects working in the same direction reinforce each other and lead to an increase in the total number of public school students at equilibrium. Vouchers, acting directly, as well as indirectly through peer quality, induce a flight of

[21] Note that under the voucher system the upper limit of the middle integral (in both numerator and denominator) is given by the cutoff ability at the pre-program public-private system, $\hat{\alpha}_0(y)$, and is not dependent on any parameter change in the voucher system. (Under the PP system, $\hat{\alpha}(y, b^e, 0, .) = \hat{\alpha}_0(y)$.)

[22] However the equilibrium may not be unique. In the presence of multiple equilibria, one cannot be sure (without a well-specified model of dynamics of adjustment) which equilibrium will be reached after a small perturbation of a parameter. To avoid these difficulties, I henceforth restrict my attention to parameter values that ensure unique equilibrium. A unique equilibrium holds if $b(0) > 0$ and $\frac{\delta g}{\delta b^e} < 1$. These conditions always hold if the marginal utility from quality ($\frac{\delta u}{\delta q}$) and the marginal responsiveness of quality to peer quality ($\frac{\delta q}{\delta b^e}$) are not very large.

high ability households at each income level at and below y_T to the private sector. The consequence is a decrease in the total number of public school students with vouchers at a household equilibrium.

An increase in the proportion β leads to a decrease in the number of public school students at a household equilibrium. This is because with an increase in β, larger number of voucher applicants are absorbed by private schools, thus decreasing the number of unsuccessful students returning to public schools. However, the marginal number of students that the public school can gain with an increase in effort increases with β. (See appendix for proof.) Under the voucher system, only β proportion of those that apply with vouchers are accepted, while $(1-\beta)$ proportion return to public school. Therefore, with a marginal increase in effort, out of the students who would like to return to public school, some are already in public school due to previous failure to get a voucher seat,—so that the number of students gained with a marginal increase in effort is smaller than if β was equal to one. If β is higher, higher proportion of voucher applicants are accepted and less students return to public school unsuccessful. Hence marginal number of students gained by a public school due to an increase in effort under higher β is larger.

4.2 Public School Behavior

The public school correctly anticipates household behavior in the future stage of the corresponding game, and chooses effort to maximize rent. Under the PP system, the revenue function is given by $pN(e,0)$ and the rent function is given by $pN(e,0) - c_1 - c(N(e,0)) - C(e)$.[23] Correspondingly, there exists a unique effort e_{PP} that solves the first order condition $\frac{\delta R(e,0)}{\delta e} = (p - c_N)N_e(e,0) - C_e(e) = 0$.[24]

Under phase I of the Milwaukee voucher system, the students participating in the choice program were included in the membership count of the MPS for revenue purposes and then the MPCP amount (obtained by multiplying the voucher amount times the number of MPCP participants) was subtracted from the MPS revenue. The revenue function under phase I of the program is therefore given by: $pN(e^*,0) - v[N(e^*,0) - N(e,v)] = pN_0 - v[N_0 - N(e,v)]$, where $N(e^*,0) = N_0$ gives the equilibrium

[23] Note that N depends on other parameters t, c, y_T, β also, but they are suppressed to simplify notation.

[24] I assume $|u_{\theta\theta}|$ is sufficiently high, that is, the rate of fall of marginal utility of quality with quality is sufficiently large. This ensures that the revenue function is strictly concave under each of the three systems (PP, voucher phase I and voucher phase II). Since the cost function is strictly convex, the rent function is strictly concave under this assumption.

number of students under the pre-program public-private system. The corresponding rent function is given by $R_{V,I}(.) = pN_0 - v[N_0 - N(e,v)] - c_1 - c(N(e,v)) - C(e)$. In phase II, the definition of membership was changed to exclude MPCP pupils from the membership count of the MPS for revenue purposes and only half the MPCP amount was funded from the MPS revenue. To incorporate this, the revenue function is modeled as $pN(e,v) - \frac{v}{2}[N_0 - N(e,v)] = pN_0 - p[N_0 - N(e,v)] - \frac{v}{2}[N_0 - N(e,v)] = pN_0 - (p+\frac{v}{2})[N_0 - N(e,v)]$. The corresponding rent function is given by $R_{V,II}(.) = pN_0 - (p+\frac{v}{2})[N_0 - N(e,v)] - c_1 - c(N(e,v)) - C(e)$. Since the voucher amount has been approximately equal to state aid per pupil, $p > v$. Denoting the loss in revenue per student due to vouchers by l, the rent functions can be denoted by:

$$R_V = (p-l)N_0 + lN(e,v) - c_1 - c(N(e,v)) - C(e), \quad \text{where}$$

$$l = \begin{cases} v & \text{in Milwaukee Phase I} \\ (p+\frac{v}{2}) & \text{in Milwaukee Phase II} \end{cases}$$

Since $(p+\frac{v}{2}) > v$, l in phase II exceeds that in phase I.[25] Let $e_{V,I}$ and $e_{V,II}$ denote the unique equilibrium efforts under the phase I and phase II programs obtained by solving the corresponding first order conditions.

Proposition 1 *Equilibrium public school effort under the voucher program can be either greater or less than the pre-program public-private equilibrium.*

In the pre-program simple public-private equilibrium, marginal revenue equals marginal cost of effort at e_{PP}. Vouchers affect both marginal revenue and marginal cost in multiple ways and these effects together determine whether or not the public school increases effort. More precisely, equilibrium effort increases iff the following expression is positive: $[(p - c_N)N_{ev} - c_{NN}N_vN_e]$ (3.2.1). Vouchers decrease the number of public school students. Since the cost function is convex in the number of students, vouchers decrease marginal cost on this account. This is captured by the second term in (3.2.1). The

[25] Even if the federal and local revenues do not go down with loss of pupils to MPCP, the loss of revenue per pupil in the second phase will be at least $(v+\frac{v}{2}) > v$. Formally, the revenue in the second phase can then be represented by: $pN_0 - (v+\frac{v}{2})[N_0 - N(e,v)] = [p - (v+\frac{v}{2})]N_0 + (v+\frac{v}{2})N(e,v)$, so the loss in revenue per pupil in the second phase is $(v+\frac{v}{2})$.

first term captures the change in net marginal revenue due to vouchers. Given that net marginal revenue per student ($p - c_N$) is positive, this depends on the effect of vouchers on the marginal number of students from a unit increase in effort (N_{ev}). This can either increase or decrease with vouchers, thus rendering the effect on public school effort ambiguous.[26] Public school effort increases if either net marginal revenue increases or the decrease in marginal revenue is less than the decrease in marginal cost.

Proposition 2 *(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.*

The first part of the proposition says that a higher revenue loss per student will lead to higher equilibrium effort. Therefore equilibrium effort under phase II of the program will be greater than phase I on this account. An increase in the revenue loss per student with vouchers implies that the revenue that can be gained by attracting a student (who would have otherwise moved to private school with vouchers) to public school is higher. A marginal increase in effort, given other parameters, attracts exactly the same number of students in both phase I and phase II and hence leads to the same increase in cost. However due to a higher l in phase II, the increase in revenue is greater in phase II. This induces public schools to supply a higher effort at equilibrium in phase II.

The second part of the proposition says that a higher proportion β would lead to a higher equilibrium effort. Therefore equilibrium effort under phase II of the program will be greater than phase I on this account. Consider two voucher systems, the only difference between them being a higher β in the second voucher system. At equilibrium under the first voucher system, marginal revenue of effort exactly equals its marginal cost. Starting from this same effort in the second system, a marginal increase in effort

[26] $N_{ev} = \int_0^1 \left[\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta v} + \frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta b} \frac{\delta b^*}{\delta v} \right] dy$. There are two effects. The first is a direct effect whereby the marginal number of students that the school can gain with a unit increase in effort falls with vouchers. Vouchers lead to an exodus of relatively high-ability households (at each income level) to private schools, so that the new marginal household (who is indifferent between the public and private sectors) has a relatively lower marginal valuation of quality. Consequently, the number of students gained due to a marginal increase in effort is lower under vouchers. This is captured by the negative first term. The second is an indirect effect. Vouchers decrease peer quality ($\frac{\delta b^*}{\delta v} < 0$) which in turn affects the marginal number of students. Since the marginal utility from school quality decreases with quality ($u_{qq} < 0$) the marginal number of students due to an increase in effort decreases with an increase in peer quality ($\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta b} < 0$). Since vouchers lead to a fall in peer quality, the marginal number of students increases due to this factor (which is captured by the positive second term).

increases the cost of effort by exactly the same amount as in the first system (C_e is same). However, as discussed in the previous section, the marginal number of students attracted under the second system is larger,—this leads to a higher marginal revenue. Also, the number of public school students under the second system is smaller, which given the convexity of cost in the number of students, leads to a lower marginal cost. Therefore the public school finds it profitable 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.*

5 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. From 1989 through 1997, school scores for this test were reported in three "performance standard categories": percentage of students below, percentage of students at, and percentage of students above the standard.[27] School scores for these three categories are available 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, 1987-97 and 1989-92 respectively. The ITBS language arts test

[27] Percentage of students below the performance standard, percentage of students at the standard, and percentage of students above the standard will be denoted by % above, % at, and % below, respectively, in the remainder of the paper.

was last administered in 1992. Starting with 1994, the ITBS was administered only in math; as of 1999, the ITBS was no longer administered as a district assessment program. The mode of reporting ITBS math scores changed in 1998. So I focus on pre-1998 ITBS math scores.

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 grade 4 took place in 1997.[28] School level grade 4 NPR scores on the five subject areas are available for the period 1997-2002.[29]

6 Empirical Strategy

The empirical part of the paper seeks to test the theoretical prediction that the quality improvement of the treated public schools in the Milwaukee voucher program will be greater in the second phase than in the first phase.

6.1 Samples and Specifications

6.1.1 Samples

I employ two alternative strategies for sample formation. Both strategies use the basic intuition in the Hoxby studies that the extent of treatment of the Milwaukee public schools depends on their pre-program percentages of free or reduced price lunch eligible students.

First Strategy: This strategy is based on Hoxby (2003a) 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

[28] In grades 8 and 10, WKCE was first administered in 1994.

[29] School scores on the WRCT are available for the second phase also. However, according to the Wisconsin DPI, the test results for 1998 and afterwards are not comparable to those in 1997 and before. The original performance standard for WRCT was established in 1989 and scores were identified into three categories based on one standard: percentage of students above, below and at standard or inconclusive. The “inconclusive” category represented a margin of error around the cut score where the scores were neither clearly above or below standard. In 1998, reporting requirements were changed to four reporting categories: minimal, basic, proficient and advanced. A new panel established new performance standards in 1998. This time three standards instead of one was used. The test format was also changed, more challenging items were included and the number of test items on the test was increased. Because of these significant changes, the DPI maintains that the pre-1998 test scores are not comparable to those at or after 1998. So 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.

treatment of the Milwaukee schools depended on the percentages of their students eligible for free or reduced price lunches.[30] Exploiting this, Hoxby classifies the Milwaukee schools into two treatment groups based on the percentages of their free or reduced price lunch students—"most treated" (Milwaukee schools where at least two-thirds of the students were eligible for free or reduced price lunches in the pre-program period) and "somewhat treated" (Milwaukee schools where less than two-thirds of the students were eligible for free or reduced price lunches in the pre-program period).

I classify the schools into three treatment groups (unlike two in Hoxby) based on their pre-program (1989-90 school year) percentages of free or reduced price lunches. So the treatment groups here are more homogenous as well as starker from each other. Also, as will be discussed below, to test the robustness of the results, I consider alternative samples that are obtained by varying the cutoffs that separate the different treatment groups.

I restrict my analysis to elementary schools only, as is Hoxby. First, as table 5 shows, most of the students participating in the MPCP were elementary grade students. This may be because the private elementary schools were less costly than private high schools (Hoxby (2003a)) and/or bulk of the participating MPCP schools offered programs in elementary grades (Wisconsin Legislative Audit Bureau Reports (1995, 2000)). Second, there were very few middle and high schools in the MPS, so that the number of schools in the different treatment groups would be too small to justify analysis.

Figure 1 shows the distribution of Milwaukee elementary public schools according to the percentages of their free or reduced-price lunch eligible students in 1990. Since schools with such population between 47% and 60% clearly form a group with an appreciable number of schools, they constitute my middle or somewhat treated group. Schools with at least 60% of their students eligible for free or reduced-price lunch are classified as "more treated" and those below 47% as "less treated". I shall denote this sample as "60-47". It consists of 42 more treated, 42 somewhat treated, and 21 less treated schools. In the more treated group an average of 82.9% of students were eligible for free or reduced-price lunch, 62.9%

[30] Under the Milwaukee program, all households at or below 175% of the poverty line are eligible to apply for vouchers. Households at or below 185% of the poverty line are eligible for free or reduced-price lunches. However the cutoff of 175% is not strictly enforced (Hoxby (2003b)) and households within this 10% margin are often allowed to apply. Also there were very few students who fell in the 175%-185% range, in fact 90% of the free/reduced price lunch eligible students qualified for free lunch. (Witte (2000)). Students below 135% of the poverty line qualify for free-lunch.

were black and 14.81% were hispanic. In the somewhat treated group an average of 53.6% were free or reduced-price lunch eligible, 50.57% were black and 3.68% were hispanic. In the less treated group an average of 37.17% were free or reduced-price lunch eligible, 45.37% were black and 3.83% were hispanic. Since it may be interesting to consider a classification where the middle (somewhat treated) group contains the mean[31] and some schools above and below the mean, I construct a second sample, the "66-47" sample.[32] To test the robustness of the results, I also consider alternative classifications, such as "66" and "60" samples, wherein schools with at least a 66% (60%) free or reduced-price lunch population are designated as more treated schools, and schools with free or reduced-price lunch population below 66% (60%) are designated as somewhat treated schools. (Note that the 66 sample corresponds to Hoxby's classification and my more treated group in the 66-47 sample corresponds to Hoxby's most treated group.) Apart from these samples, I have also experimented with other samples such as 75-47, 75, and 50 samples which are defined similarly. The results for these samples are broadly similar to the above four and hence will not be reported here.

The control group criteria used here is based on Hoxby (2003a). Since all schools in Milwaukee were potentially affected by the program, she constructs a control group that consists of Wisconsin schools outside Milwaukee that satisfy the following criteria: (i) they were urban (ii) had at least 25% of their population eligible for free or reduced-price lunch, and (iii) had black students compose at least 15% of the school population. Her control group consists of 12 schools.

I follow Hoxby in my control group classification, although there is some difference. The 1989-90 Public Elementary/Secondary School Universe Data Set of the NCES Common Core of Data classifies schools into seven locales (1-7) according to their locations. All Milwaukee schools fall in either of two categories—1 (large central city) and 3 (urban fringe of large central city). I picked elementary schools outside Milwaukee but within Wisconsin that in 1989-90 had at least 25% of their students eligible for

[31] The mean percentage of free or reduced-price lunch students in the Milwaukee Public Schools in 1990 was 59%.

[32] Here schools with a free or reduced-price lunch population between 47% and 66% form the somewhat treated group; those with at least 66% such population form the more treated group; those below 47% form the less treated group. It contains 33 more treated, 53 somewhat treated and 21 less treated schools. Under this classification, the more treated group has an average of 84.5% free or reduced-price lunch eligible students, 66.5% black, and 18.07% hispanic students. The somewhat treated group has an average of 55.4% free or reduced-price lunch eligible students, 50.99% black, and 4.09% hispanic students.

free or reduced price lunches, had black students constitute at least 15% of their school population and were as similar as possible to the Milwaukee schools in terms of their locales to form my control group. No elementary school outside Milwaukee (but within Wisconsin) had a locale code of 1. Therefore I picked elementary schools that satisfied the above two criteria in terms of their black and free/reduced price lunch populations and had a locale code of 2 (middle-size central city), 3 or 4 (urban fringe of mid-size city). The control group thus constructed contained 33 schools. Most of these schools had a locale code of 2 and very few had codes 3 and 4. These schools come from four school districts—Beloit, Kenosha, Madison Metropolitan and Racine. Geographically also, they are located close to Milwaukee.[33] In this untreated comparison group, an average of 44.95% of the students were eligible for free or reduced-price lunches, 22.37% were black, and 14.84% were hispanic. This group forms the control group for each of the above samples of treated schools. The groups thus constructed will form my treatment and control groups for the first phase as well as the second phase of the program, that is, the schools in the different treatment and control groups remain exactly the same in the analyses of the first and second phases. Using each of these samples, I investigate how the different treatment groups in Milwaukee responded to the voucher program in the first phase and in the second phase of the program in comparison to the control group of schools.

Second Strategy: 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 is a wide variation across Milwaukee schools in the percentage of their free or reduced-price lunch students. In 1990, some schools had as few as 22% of their students eligible for free or reduced-price lunches, while others had as large as 93% of their students eligible. Exploiting this variation, I

[33] Milwaukee, Kenosha and Racine are located in CESA 1 (Cooperative Educational Service Agency). Beloit and Madison are located in CESA 2 which borders CESA 1. The state of Wisconsin is organized into 12 CESAs.

investigate whether an increase in the intensity of treatment is 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.

6.1.2 Specifications

This study considers public school scores as the outcome variable. The initial program was announced in 1990 and was first implemented in the 1990-91 school year. As outlined earlier, the program saw a major shift in 1998 which took effect in the 1998-99 school year. The 1990 and 1998 programs can be looked upon as two separate programs, more so because the shift took place eight years after the initial program. (In spite of that, the regressions investigating the second phase improvement control for pre-program trends, so that any long term effects of the initial program are differenced out.)

First consider the estimation strategy corresponding to the treatment-control group classification. The identifying assumption in the analysis of both phases is that if the different treatment and control groups have similar trends in the corresponding pre-program period, any shift of the treated groups in comparison to the control group in the post-program period can be attributed to the program. To test the identifying assumption, I first run the following fixed effects regression (and also the OLS counterpart of it) using only pre-reform or pre-program data. Pre-program data for the first phase span 1987-90 while those for the second phase span 1997-98.

$$s_{it} = f_i + \beta_0 t + \beta_{1,MT}(MT * t) + \beta_{1,ST}(ST * t) + \beta_{1,LT}(LT * t) + \gamma X_{it} + \epsilon_{it}$$

where s_{it} is ith school score in year t, f_i are school fixed effects, t denotes time trend, MT is a dummy variable taking a value of 1 for more treated schools and 0 otherwise, ST is a dummy variable taking a value of 1 for somewhat treated schools and 0 otherwise, LT is a dummy variable taking a value of 1 for less treated schools and 0 otherwise, $(MT * t)$, $(ST * t)$ and $(LT * t)$ are interactions between trend and MT, ST and LT respectively, X_{it} denotes the set of school characteristics and ϵ_{it} is a stochastic error term. $\beta_{1,MT}$, $\beta_{1,ST}$ and $\beta_{1,LT}$ capture the pre-program differences in trend of the MT, ST and LT schools in comparison to the control schools.[34]

[34] When there are data on more than two pre-program years (as in the case of ITBS reading and Math in the first

If the treated and control groups have the same pre-program trend, 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. I begin with a completely linear model:

$$s_{it} = f_i + \alpha_0 t + \alpha_1 v + \sum_I \alpha_{2I}(I * v) + \alpha_3(v * t) + \sum_I \alpha_{4I}(I * v * t) + \alpha_5 X_{it} + \epsilon_{it}, \quad \text{where } I = \{MT, ST, LT\} \tag{1}$$

Here v is the program dummy, $v = 1$ if year > 1998 and 0 otherwise. The specifications shown here are for the second phase. The first phase specifications are the same except that the years are different. (For example, for the first phase, $v = 1$ if year > 1990 and 0 otherwise.) The variables v and $v * t$ respectively control for post-program common intercept and trend shifts such as national, state and county level shifts. The coefficients on the interaction terms $(I * v)$ and $(I * v * t)$, $I = \{MT, ST, LT\}$ estimate the program effects—$\alpha_{2,I}$, capture the intercept shifts and $\alpha_{4,I}$ the trend shifts for the MT, ST and LT schools respectively compared to the control group of schools. Note that one would expect the effects to have a strict hierarchy—MT effects should exceed the corresponding ST effects and the ST effects the corresponding LT effects. All specifications I describe here are fixed effects regressions. I also estimate OLS counterparts of each of these specifications. All OLS regressions include dummies for the different treatment groups.

The second model allows the trend in the comparison group to be non-linear while still constraining the year-to-year gains of the treated schools in the post-program period to be linear in addition to an intercept shift.

$$s_{it} = f_i + \sum_{i=1998}^{2002} \beta_i D_i + \sum_I \beta_{0I}(I * v) + \sum_I \beta_{1I}(I * v * t) + \beta_2 X_{it} + \epsilon_{it}, \text{ where } I = \{MT, ST, LT\} \tag{2}$$

phase) I also fit a non-linear specification with pre-program year dummies and interactions of the treated dummies with pre-program year dummies. This allows the individual pre-program year effects of the treated schools to vary in an unrestricted way from those of the control schools.

where D_i, $i = \{1998, .., 2002\}$ are year dummies for 1998 through 2002 respectively. β_{0I} and β_{1I}, $I = \{MT, ST, LT\}$ reflect the post-program intercept and trend shifts respectively of the MT, ST and LT groups after controlling for common post-program year effects.

Finally, I estimate a completely unrestricted and non-linear model that includes year dummies to control for common year effects and interactions of post-program year dummies with the MT, ST and LT dummies respectively to capture individual post-program year effects.

$$s_{it} = f_i + \sum_{i=1998}^{2002} \gamma_i D_i + \sum_{i=1998}^{2002} \gamma_{0i}(MT * D_i) + \sum_{i=1998}^{2002} \gamma_{1i}(ST * D_i) + \sum_{i=1998}^{2002} \gamma_{2i}(LT * D_i) + \gamma_3 X_{it} + \epsilon_{it} \tag{3}$$

This specification no longer constrains the post-program year-to-year gains to be equal and allows the program effect to vary across the different years. The coefficients γ_{0I} γ_{1I} and γ_{2I}, $I = \{1998, .., 2002\}$ represent respectively the more treated, somewhat treated and less treated year effects after one, two, three and four years into the program.

Now consider the estimation strategy corresponding to the continuous variable formulation. For each of the two phases, I estimate versions of the above three regressions after appropriately adjusting them for a continuous variable:

$$s_{it} = f_i + \delta_0 t + \delta_1 v + \delta_2(\%frl * v) + \delta_3(v * t) + \delta_4(\%frl * v * t) + \delta_5 X_{it} + \epsilon_{it}$$
$$s_{it} = f_i + \sum_{i=1998}^{2002} \phi_i D_i + \phi_0(\%frl * v) + \phi_1(\%frl * v * t) + \phi_3 X_{it} + \epsilon_{it}$$
$$s_{it} = f_i + \sum_{i=1998}^{2002} \psi_i D_i + \sum_{i=1999}^{2002} \psi_{1,i}(\%frl * D_i) + \psi_3 X_{it} + \epsilon_{it}$$

The corresponding OLS regressions also include the variable $\%frl$. The linear and semi-linear specifications investigate whether intercept or trend shifts are associated with an increment of treatment intensity, while the non-linear specification investigates whether an increment in treatment intensity is associated with year effects in the first, second, third and fourth years after program. It may be noted here that, in this approach, the program effect is assumed to vary linearly with treatment intensity. To relax this assumption, I also estimate modified versions of the above specifications. The linear specification then includes interactions between treatment intensity, dummy variables representing different treatment groups and program dummy ($\%frl * I * v$ where $I = \{MT, ST, LT\}$) and interactions between treatment intensity, treatment group dummy variables, program dummy and trend ($\%frl * I * v * t$ where $I = \{MT, ST, LT\}$). The non-linear specification includes interactions between treatment intensity, group dummy variables and year dummies($\%frl * I * D_i$). The treatment groups used here are the more treated, somewhat treated and less treated groups described above. (Here, I consider treatment group classification corresponding to each of the samples above.)

6.2 Mean Reversion

However, there are several concerns that are worth considering. I discuss these and their potential concerns one by one. First is the issue of mean reversion. Mean reversion is the statistical tendency whereby high or low scoring schools tend to score closer to the mean subsequently. The more treated schools were among the lowest scoring schools in each of the subject areas in both 1990 and 1998—in particular, the more treated average scores in each of the subject areas were lower than each of the corresponding somewhat treated, less treated and control school average scores, and in many cases, these differences were statistically significant. Therefore, a natural question to ask would be whether the improvement (if any) in the Milwaukee program is driven by mean reversion rather than the program. Since I do a difference-in-differences analysis, my estimates would be contaminated if the more treated schools exhibit mean reversion in comparison to the control group of schools. Note that mean reversion can be a problem in the estimation corresponding to the treatment-control group strategy only, but not corresponding to the continuous variable estimation strategy.

To address the issue of mean reversion, I use the pre-program data for the corresponding phase. To investigate the issue of mean reversion in the second phase, I 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. To implement this strategy, I use two alternative methods. In method 1, I first construct an index which is the sum of NPR scores in the five subjects of reading, language arts, math, science and social studies. Based on this index, I rank the Milwaukee schools in

1998 and note the ranks of each of the more treated, somewhat treated and less treated schools. Then using 1997 WKCE scores, I rank the schools in terms of this index in 1997. Using the 1997 ranks, I pick schools in 1997 that have exactly the same ranks as the more treated schools in 1998. I call this group of schools, the "low" group. Similarly, using the ranks of the somewhat treated (less treated) schools in 1998, I pick schools in 1997 that have the same rank as the 1998 somewhat treated (less treated) schools, and call them the "mid" ("high") schools. The intuition here is that any improvement of the "low" schools in comparison to the control schools[35] during the period 1997-98 can be characterized as the mean reversion effect of the more treated schools as this was before the program shift. I then subtract out this mean reversion effect from the program effect of more treated schools in phase II to arrive at the mean reversion corrected effect.

In method 2, I rank the Milwaukee schools in 1998 on the basis of NPR scores of each subject, and calculate the mean reversion effect of each subject solely based on ranks of schools in that subject. For example, ranking schools in 1998 in terms of NPR reading scores, I note the ranks of the more treated, somewhat treated and less treated schools. Then I rank the schools in 1997 based on their 1997 WKCE reading NPR scores and pick schools that have the same rank as the more treated in 1998 and call then the "low" group. Similarly, I construct the "mid" and "high" groups in 1997. If the "low" group thus constructed exhibit an improvement relative to the control schools in reading during 1997-98, I call this the mean reversion effect in reading and subtract it out from the more treated program effect in reading obtained earlier to arrive at the mean reversion corrected effect in reading. Similarly, based on ranks of schools in each of the other subjects in 1997 and 1998, I calculate the mean reversion corrected effect in the corresponding subjects. To investigate the issue of mean reversion in phase I, I use exactly the same two methods as above except that the years 1997 and 1998 are respectively replaced by 1989 and 1990 and schools are ranked based on test scores available in that period (WRCT and ITBS) rather than WKCE.

[35] Controls schools are the same as earlier.

6.3 Competitive Effect–Presence of Voucher Schools

The second 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 was a strong private school presence. More specifically, even schools that had a major proportion of their students eligible for vouchers (example, the more treated schools), would not be induced to improve if there were not enough voucher schools in close proximity to absorb them.

To investigate this issue, I examine the distribution of voucher schools in Milwaukee and their distances from public schools.[36] 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. Table 3 shows the distribution of voucher schools around Milwaukee schools. 27% of the public schools had 1-2 voucher schools within a one mile radius, more than 19% had 3-5, 30% had 6-10 and 13% had more than 11 voucher schools within a one-mile radius. The presence of voucher schools around the more treated schools was even stronger. 23% of the more treated schools had 3-5 choice schools within a one mile radius, 48% had 6-10 and 26% had more than 10. As table 3 shows, this picture is mirrored in the distribution of choice schools within 2 and 3 mile radii of an average Milwaukee school and an average more treated school. Thus there is strong evidence of considerable choice school presence, more so in the vicinity of more treated schools. This confirms that the more treated schools were the ones who 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.

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

I also check whether a higher proportion of student eligibility is associated with a corresponding higher loss of voucher students from more treated schools. Table 4 shows the distribution of students lost due to vouchers from more treated, somewhat treated and less treated schools. For ease of comparison, I

[36] Public and voucher school addresses obtained from the Wisconsin DPI are used for the computation of distances.

have normalized the numbers in each year in terms of the students lost by the less treated schools in that year. In 1999, a typical more treated (somewhat treated) school lost 1.43 (1.37) times the number of students lost by the corresponding less treated school. In 2000, a more treated school lost more than twice while a somewhat treated school lost almost 1.5 times the number lost by a typical less treated school. The picture is similar in 2001 and 2002. The losses have the expected hierarchy in the sense that the more treated loss always exceeded the somewhat treated loss and the somewhat treated loss the corresponding less treated loss. Moreover in many cases the group losses are statistically different from each other. Therefore, the table corroborates the fact that higher intensity of treatment was associated with higher loss of voucher students and higher competitive pressure.

6.5 Sorting

Another issue relates to 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.[37]

To consider this issue, the following points may be noted. First, the empirical part of the paper seeks to test the theoretical prediction that the *quality* under Milwaukee phase II will exceed that under Milwaukee phase I (Corollary 1), where quality is a combination of public school effort and peer quality, so there is a one to one correspondence between the theory and empirics. Second, each of the regressions control for demographic composition of schools (example, racial and sex compositions of schools and percentage of students eligible for free or reduced price lunches). However any change in student composition in terms of unobservable factors may not be controlled for by these variables. Note that if sorting leads to cream-skimming, this is going to 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. Therefore inability to adequately account for sorting in this case would only lead to an underestimate of the differential second phase effect (as compared to the first phase).

Finally, to investigate this issue, I examine whether the demographic composition of the different

[37] See Hsieh and Urquiola (2003) for a discussion.

Milwaukee treated groups changed over the years. I run the same three specifications as above except that the dependent variable of school scores is replaced by the respective demographic variable (% black, % hispanic etc.). I do not find much evidence of changes in demographic compositions of schools, either in phase I or in phase II (table 6). Only a few of the coefficients are statistically significant and they are always very small. In phase I, the coefficients suggest trend shifts of less than 1% while in phase II, these shifts range between 0.8%-1.38%. This provides suggestive evidence that sorting was not an important factor.

7 Results

Milwaukee Phase I

Table 7 compares the pre-program trends of the more treated, somewhat treated and less treated schools with that of the control schools in reading, math, and writing. The odd-numbered columns present OLS estimates with standard errors that allow for correlations within districts. The even-numbered columns present fixed-effects estimates. The table shows no statistically significant evidence of any difference in pre-program trend between the different groups.

Using the 66-47 sample, table 8 analyzes the effect of the Milwaukee voucher program in phase I on WRCT (% above and % below) scores and ITBS reading, math and language arts scores. For each set, the first column reports results from the linear model and the second from the non-linear model. All results reported are from regressions that include school fixed effects. Results from OLS regressions as well as specification (2) are similar and hence skipped. (They are available on request.) For WRCT (both % above and % below), the columns (2) and (4) show positive and statistically significant effects in the second and fourth years after program. 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.[38] Although many of the other effects are positive (both in linear and non-linear specifications),

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

they are not significant and do not always have the right hierarchy between groups.

The results for ITBS reading show no evidence of any statistically significant effect of the program on any of the treated groups and the effects do not always have the right hierarchy. However, since many of the effects are positive and non-negligible in magnitude, I also conduct an F-test of joint significance of the more treated effects. They are never statistically significant for ITBS reading. (Note that the more treated effects are jointly significant for WRCT (% above) and in the non-linear specification for WRCT (% below).) The picture for ITBS math is broadly similar, except that the second year effect of the somewhat treated group is positive and significant, although the corresponding more treated effect is smaller. In language arts, both the linear and non-linear models show a deterioration immediately after the program which is statistically significant in the case of the more treated schools. This deterioration is however temporary, and is reversed in the second year after program. Once again, the somewhat treated effects, exceed the more treated effects economically, though never statistically.[39] The F-test for joint significance of the more treated effects show no evidence of any effect for ITBS math and a deterioration in ITBS language arts. Figure 2 graphs the OLS estimates for ITBS from the linear model. As expected there is no evidence of any program effect in reading and math and a deterioration in language arts in its first year.

Table 9 considers a continuous treatment variable and proxies the intensity of treatment by the pre-program (1990) percentage of free or reduced price lunch eligible students of schools. The first column of each set presents estimates from the linear specification while the second from the non-linear specification. Although once again the second and fourth year effects in WRCT (both % above and % below) provide some evidence of improvement economically, they are no longer 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 its first year. Moreover, in neither of the regressions are the treatment effects jointly significant (except in language arts). As discussed in the empirical strategy

[39] 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. (Although some other districts in Wisconsin also administered the ITBS, they often used other forms of the test. The modes of reporting scores were also different between different districts and hence not comparable.) 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.

section, this assumes the program effect to vary linearly with treatment intensity. Therefore, I also run regressions that allow the program effect to vary non-linearly with treatment intensity (see section 6.1). Since these results are qualitatively similar to those above, they are not reported here.

The findings for phase I can be summarized as follows: 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. There is some evidence of a positive effect in the second and fourth years after program in WRCT, at least in the treatment group analysis. There is no evidence of any effect in ITBS reading, math and language arts (except for some evidence of an initial deterioration in language arts.) These results seem to be robust in that they are replicated in the analysis with other samples for each of the above tests.[40] However, the effects for the different treated groups are never statistically different from each other,—not even for the WRCT second and fourth year estimates.

Milwaukee Phase II

Investigation of pre-reform trends using WKCE data for 1997 and 1998 reveals that there is no differential trend in reading, language arts and science between the different groups of schools. However, the more treated schools exhibit a positive significant differential trend in comparison to the control group of schools in both math and social studies. (These results are not reported for lack of space but are available on request.) Using the 66-47 sample, table 11 analyzes the effect of the Milwaukee voucher program on WKCE reading, language arts, math, science and social studies scores in the second phase of the program. For each set, the first column presents results from the linear model and the second from the non-linear model. Whenever there are differences in pre-program trends, the results reported control for these differences.[41]

The results for reading and language arts are similar. Estimation of the linear model shows positive intercept and trend shifts in most cases, although they are not always statistically significant. The nonlinear model estimations (columns (2) and (4)) yield positive year effects which are statistically significant in most cases. Moreover, the effects (both in the linear and non-linear specifications) almost

[40] The results for some other samples for WRCT are illustrated in appendix table A.1.

[41] For each of the subject areas, I have estimated regressions that control for pre-program trends as well as those that do not. The results are qualitatively similar under both formulations.

always have the right hierarchy and are often statistically different between groups.

In math, science and social studies, although the effects from the linear model are in most cases positive, they are often not statistically significant and do not have the expected hierarchy. The estimates from the non-linear model in math and social studies (columns (6) and (9)) are positive and statistically significant in most cases. However, the effects often do not have the right hierarchy. The results from the nonlinear model in science show positive and significant effects, at least for the somewhat treated and more treated effects. However, the more treated effect is larger than the somewhat treated effect only in the last year after program, though the effects are not statistically different from each other.

Figure 3 graphs the OLS estimates from the linear model for reading, language arts and math. A vertical line is drawn at 1998 to characterize the program shift. Consistent with the results above, they 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/time path 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.

The results are quite robust in that for each subject area, the same set of findings hold for different samples,[42] different specifications and both OLS and FE estimates for each specification. These findings can be summarized as follows. In both reading and language arts, there is considerable evidence of improvement in phase II of the program. Moreover, the more treated effects exceed the somewhat treated effects and the somewhat treated effects exceed the corresponding less treated effects in most cases.[43] In WKCE science, the initial improvement of the somewhat treated group exceeds that of the more treated group, however over the years the more treated group has improved at a higher rate than

[42] The results for WKCE reading, language arts, math and social studies for other samples such as 60-47, 66, 60 etc. are very similar to those described above. Results for some of the subject areas for these samples are reported in appendix table A.1.

[43] Interestingly, the effects are considerably larger in the second, third and fourth years after program than in the first year. It may be remembered in this context that the first year effect will be solely due to the expansion of the program to include religious schools, whereas the next three years would capture the effect of both this expansion as well as an increase in the monetary loss from vouchers. However, this does not imply that the latter policy is more effective than the former because most programs take time to generate desired effects and it is often easier to respond gradually to a program.

the somewhat treated group, so much so that the fourth year effect surpasses 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.

Table 10 considers a continuous treatment variable and proxies the intensity of treatment by the pre-program (1990) percentage of free or reduced-price lunch eligible students of schools. It investigates whether an increase in treatment intensity is associated with an improvement in WKCE reading, language arts, math, science and social studies scores. The first column of each set fits a completely linear model—after controlling for post-program common intercept and trend shifts and any pre-program difference in trend, it investigates whether an increase in treatment intensity is associated with an intercept and/or a trend shift. The second column of each set fits a nonlinear model. After controlling for any pre-program difference in year effects and common post-program shocks (using year dummies), it investigates whether an increase in treatment intensity is associated with an improvement in the first, second, third and fourth years after program. There is 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. Thus the results from this table are consistent with those obtained in table 11.

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.[44] 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 the 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. Traditionally, teachers were hired on the basis of seniority only. A

[44] There is not much anecdotal evidence of improvement in the first phase. Howard Fuller, the superintendent of Milwaukee Public Schools during 1991-95, writes: "...during its (MPCP's) early years, I observed only a limited impact on the MPS."

teacher-evaluation system was established that had union members weeding out bad teachers.[45]

Finally, it might be useful to compare the results obtained above for the treatment group analysis with the corresponding effects in Hoxby. In Hoxby (2003b)[46], 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 in table 11 (columns (4), (6) and (8) respectively) shows that the effects are qualitatively very similar. The differences in the actual magnitudes can be attributed to the following reasons. First, her effects are average annual effects while this paper considers the effects separately over the years. Second, the control group of schools here is somewhat different from Hoxby (see section 6.1.1). Third, the regressions here control for demographic characteristics and real per pupil expenditure, unlike the relevant regressions in Hoxby. (Hoxby reports separate regressions that check for the fact that the improvement is not due to increases in per pupil spending. In these regressions, productivity is defined as NPR scores per thousand dollars of per pupil spending. She finds that productivity in language, math and science rose by 0.902, 0.973 and 1.660 NPR points per thousand dollars in the more treated schools, after the reform. These results are not directly comparable to mine.)

Milwaukee Phase I Versus Phase II

Table 12 compares the effect of the voucher program in Milwaukee phases I and II using both treatment group analysis (Panel A) and continuous treatment variable analysis (panel B). The estimates here are based on non-linear model estimates in tables 8-11 and all figures are expressed in terms of the respective sample standard deviations. (The results from the other models are similar and hence are not reported here.) Columns (1)-(4) report the comparison results in reading, columns (5)-(6) in language arts and columns (7)-(8) in math.

First, consider panel A. In phase I reading, both for WRCT (% above) and WRCT (% below) scores, there is statistically significant evidence of improvement in the second and fourth years after program.

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

[46] I consider Hoxby (2003b) because similar to my phase II analysis, the post-reform period considered by the study is 1998-2002. 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.

In ITBS, although the effects are positive, they are never statistically different from zero. In phase II, on the other hand, there is positive significant evidence of improvement in WKCE reading in each of the second, third and fourth years after program—and each of the Milwaukee phase II effects exceed the corresponding phase I effects for both WRCT and ITBS. In language arts, there is no evidence of improvement in phase I. On the other hand, there is positive, significant and large 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 the second, third and fourth years after program—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.[47]

The comparison results in Panel B are very similar.[48] Once again the phase II effects in reading, math and language arts exceed the corresponding phase I effects in each of the years (except first year ITBS reading, which however is not statistically significant.) 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.[49] This finding is quite robust since it holds for all the different samples, different specifications and different tests. These findings support the prediction obtained from theory.

Mean Reversion

As discussed earlier, a potential concern here is that the improvement of the more treated group can,

[47] These findings hold for each of the other samples 60-47, 66 and 60. (This can be checked for reading using appendix table A.1. The results from ITBS for samples 60-47, 66 and 60 are not reported here to save space. They are available on request.)

[48] The only difference is that the phase II effects in math and phase I second and fourth year reading effects are no longer statistically significant.

[49] Since many of the coefficients in Milwaukee phase I are not significant though positive (or negative for WRCT (% below)), 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 $\frac{1}{2}$. In each of panel A and panel B, I have 17 pair-wise comparisons. Under the null, $D =$(Milwaukee phase II effect-Milwaukee phase I effect) follows a binomial $(17, 0.5)$ distribution. D is positive in all 17 cases in panel A and 16 cases in panel B. The probability of getting 17 (16) positive D under the null is $(0.5)^{17}$ $((0.5)^{16})$. Since this is very small, the null of equal effects can be comfortably rejected.

to some extent, be caused by mean reversion. Note that this concern pertains only to the treatment group analysis and not to the continuous variable analysis. Using data for 1997 and 1998, table 13 presents the mean reversion effect estimates. Panel A uses method 1 while panel B uses method 2. There is no evidence of any mean reversion in reading, language arts or science. Both "low" and "mid" groups show comparable amounts of mean reversion in math and only "low" group shows mean reversion in social studies in the fixed effects estimate. The mean reversion results for method 2 are presented in panel B. The results are very similar except that the "low" group also shows mean reversion in science in addition to math and social studies.[50]

Table 14 compares the impact of the Milwaukee voucher program in phase I and phase II after correcting for mean reversion. The phase II effect sizes are obtained by subtracting the effect size attributed to mean reversion (obtained from expressing the relevant coefficients in table 13, in terms of standard deviations) from the more treated effect sizes (reported in table 12 panel A) in each of the four years after program. The phase I effect sizes are the same as in table 12 panel A (see footnote 50). Table 14 Panel A reports mean reversion corrected estimates obtained using method 1 while panel B reports those obtained using method 2. In both panels, the effect sizes for phase II reading and language arts remain the same as earlier. In math, although the effect sizes fall, they are still positive and considerably larger than those in Milwaukee phase I. These results strengthen my earlier findings and further confirm the validity of the theoretical prediction.

7.1 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. Charter schools have been allowed to enter in Milwaukee from the 1993-94 school year. Upto 1998, the only chartering agency in Milwaukee was the MPS and this resulted in only one charter school. This slow

[50] It may be noted here that using this strategy and the two alternative methods, I find no evidence of mean reversion in the period just before 1990 (that is, the period just preceding phase I). Hence the results from this analysis are not reported here.

growth of charters led the Wisconsin State Legislature to authorize three other chartering agencies—the city of Milwaukee, University of Wisconsin-Milwaukee and the Milwaukee Area Technical College (MATC). Starting from 1998-99, they could also grant charters in addition to the MPS. Although the growth of charter schools was initially slow even after 1998 (3 schools and 186 students in 1998-99, 5 schools and 1,239 students in 1999-2000), it 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 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, 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 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 were similar to those of an average Milwaukee school. An average Milwaukee school has 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 in the non-linear regressions for treatment group analysis do not change results qualitatively.

PAVE and Chapter 220 Programs

Two other choice programs in Milwaukee are worth mentioning and it is important to rule them out as explanations for the pattern of results obtained. Chapter 220 Program, established in 1978 and further expanded in 1987, caters to the goal of metropolitan integration. It allows minority students from the MPS to attend public schools in the twenty four suburban districts, while white students from the suburbs may enroll in the MPS. The voucher program effects in Milwaukee are not likely to be contaminated by this program since this program started much before the MPCP,—controlling for differences in pre-program trends between treatment and control schools gets rid of any effect of the Chapter 220 program, more so because the size of the latter program was relatively stable upto the late 1990s. After that, it actually shrank in size. Thus the chapter 220 program is not likely to lead to overestimates of the voucher effect in either phase I or phase II.

The PAVE (Partners Advancing Values in Education) program was established in 1992 and it came into operation from the 1992-93 school year. This is a privately funded school choice program that allows students at or below 185% (revised to 175% in 1995-96) of the poverty line in the city of Milwaukee (not just the MPS) to attend any private school in Milwaukee. Unlike the MPCP, PAVE covers only one-half of the private school tuition requiring the parents to match the other half. Although the initial participation in PAVE was not negligible, it petered out after the expansion of the Milwaukee program in 1998 and currently stands at approximately 700 students per year. Also, the proportion of students transferring from the MPS is small, always constituting less than one-third of the total PAVE population, so that the number of students leaving the MPS under PAVE has always been much smaller than under MPCP (even in the first phase). Moreover since PAVE required the scholarship to be topped up, overwhelmingly white and more advantaged households participated in the PAVE and the demographic composition of the PAVE students differed substantially from that of the MPCP students. The more-treated schools in this paper are predominantly black and hence are not likely to be strongly affected by PAVE. Further, there is no evidence of any trend shift in scores of the different treatment groups in 1992-93, the first year after PAVE. Finally, if anything PAVE will lead to overestimates of the first phase effect, but not the second phase when students leaving the MPS due to PAVE was very small. This would indicate even larger (rather than smaller) improvement differences between the first and second phases than that indicated in the paper.

Accountability System

Wisconsin had an accountability system in place from 1997-98. 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. Thus, this factor is unlikely to bias the program effect.

8 Conclusion

This paper analyzes the role of vouchers as instruments of public school reform. It presents strong evidence that voucher design matters. It shows both theoretically and empirically 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 parental choice program. The second phase saw a major increase in the number of private school seats when following a Wisconsin Supreme court ruling religious schools were allowed to participate for the first time in the 1998-99 school year. The second phase was also characterized by a discontinuous rise in the revenue loss per student from vouchers due to some changes in the funding formula. In the context of an equilibrium theory of public school and household behavior, the study predicts that these factors would lead to an unambiguous improvement in public school performance. Using a difference-in-differences analysis in trends and Wisconsin data from 1987 through 2002, it then shows that this prediction is validated empirically. The paper thus provides an important lesson—any voucher program may not have a positive effect on public school incentives and performance. However, careful choice of parameters can go a long way to induce public school improvement. The findings of the paper undoubtedly have important implications for public school reform, more so in the context of the present concern over public school performance.

This paper considers public school test scores as the only outcome variable. An interesting direction of future research would be to analyze the impact of such policy changes on other outcome variables—such as absentee rate, retention rate, dropout rate, teacher quality etc. Presumably increase in competition (through changes in parameters such as above) would induce public schools to make themselves more attractive to their potential customers and to cater to the amenities that their customers care for. This relates to the deeper question of the household preference function and what households value most in public schools.

Technical Appendix: Proofs of Results and Claims

Claim 1: *A household equilibrium always exists.*

The existence can be proved in the following steps:

(i) Define $\Phi : [0,1] \to [0,1]$ such that for all $b' \in [0,1]$,

$$b = \Phi(b') = \frac{\int_0^{y_T} \int_0^{\hat{\alpha}(y,b',v,.)} \alpha d\alpha dy + (1-\beta) \int_0^{y_T} \int_{\hat{\alpha}(y,b',v,.)}^{\hat{\alpha}_0(y)} \alpha d\alpha dy + \int_{y_T}^1 \int_0^{\hat{\alpha}(y,b',0,.)} \alpha d\alpha dy}{\int_0^{y_T} \int_0^{\hat{\alpha}(y,b',v,.)} d\alpha dy + (1-\beta) \int_0^{y_T} \int_{\hat{\alpha}(y,b',v,.)}^{\hat{\alpha}_0(y)} d\alpha dy + \int_{y_T}^1 \int_0^{\hat{\alpha}(y,b',0,.)} d\alpha dy}$$

Define a function F such that $F(\hat{\alpha}(.), y_T) = \frac{\int_0^{y_T} \int_0^{\hat{\alpha}(v,.)} \alpha d\alpha dy + (1-\beta) \int_0^{y_T} \int_{\hat{\alpha}(v,.)}^{\hat{\alpha}_0} \alpha d\alpha dy + \int_{y_T}^1 \int_0^{\hat{\alpha}(0,.)} \alpha d\alpha dy}{\int_0^{y_T} \int_0^{\hat{\alpha}(v,.)} d\alpha dy + (1-\beta) \int_0^{y_T} \int_{\hat{\alpha}(v,.)}^{\hat{\alpha}_0(y)} d\alpha dy + \int_{y_T}^1 \int_0^{\hat{\alpha}(0,.)} d\alpha dy}$.

(ii) $\hat{\alpha}(y, b', .)$ is continuous in b' (from 3.1.1a). $F(\hat{\alpha})$ is continuous in $\hat{\alpha}$ (as both numerator and denominator are continuous in $\hat{\alpha}$ and $0 < \hat{\alpha} < 1$ ensures that the denominator is non-zero). Therefore Φ is a continuous function from $[0,1] \to [0,1]$.

(iii) Since $[0,1]$ is non-empty, compact and convex and Φ is continuous, there exists at least one fixed point $b^* = \Phi(b^*)$ by Brouwer's fixed point theorem.

Claim 2: *The equilibrium number of public school students increases with public school effort and decreases with vouchers.*

Proof:

Step 1: *Equilibrium peer group quality falls with vouchers and increases with public school effort.*

At the household equilibrium,

$b^* = g(b^*, e, v, t, c, y_T, \beta)$ where b^* denotes the equilibrium peer quality under targeted vouchers.

Effect of an increase in e:

$$\frac{\delta b^*}{\delta e} = \frac{\frac{\delta g(b^*,.)}{\delta e}}{1 - \frac{\delta g(b^*,.)}{\delta b}}$$

The denominator is positive from the uniqueness condition. Consider the numerator.

$$\frac{\delta g(b^*,.)}{\delta e} == \frac{1}{N(b^*,.)}\left[\beta y_T(\hat{\alpha}(v,b^*,.)-b^*).\frac{\delta\hat{\alpha}(v,b^*,.)}{\delta e} + (1-y_T)(\hat{\alpha}(0)-b^*)\frac{\delta\hat{\alpha}(0,b^*,.)}{\delta e}\right]$$

Using equation 3.1.1, $\frac{\delta^2\hat{\alpha}}{\delta v\delta e} = \frac{u_e^q.\frac{\delta\hat{\alpha}}{\delta v}}{u_\alpha^Q - u_\alpha^q}$, which is negative.

Consider $\int_0^1 \hat{\alpha}dy - b = y_T[\hat{\alpha}(v,.) - \frac{\int_0^{\hat{\alpha}(v,.)}\alpha d\alpha}{N} - (1-\beta)\frac{\int_{\hat{\alpha}(v,.)}^{\hat{\alpha}(0,.)}\alpha d\alpha}{N}] + (1-y_T)[\hat{\alpha}(0,.) - \frac{\int_0^{\hat{\alpha}(0,.)}\alpha d\alpha}{N}]$. Note that $\hat{\alpha}(.) > \frac{\int_0^{\hat{\alpha}(.)}\alpha d\alpha}{\int_0^{\hat{\alpha}(.)}d\alpha} > \frac{\int_0^{\hat{\alpha}(.)}\alpha d\alpha}{N(.)}$, for each y. $\int_0^1 \hat{\alpha}dy - b$ is positive if β is not very small. Henceforth, I assume β is not very small—in Milwaukee, its value was quite large (on average, 0.78 during 1991-95). Then $\int_0^1 \hat{\alpha}dy - b > 0$. Therefore,

$$\int_0^1 \hat{\alpha}(.)dy - b > 0 \Rightarrow y_T\hat{\alpha}(v,.) + (1-y_T)\hat{\alpha}(0,.) - b > 0 \Rightarrow y_T[\hat{\alpha}(v,.)-b] + (1-y_T)[\hat{\alpha}(0,.)-b] > 0$$

$$\Rightarrow (1-y_T)[\hat{\alpha}(0,.)-b] > |y_T[\hat{\alpha}(v,.)-b]| \Rightarrow (1-y_T)[\hat{\alpha}(0,.)-b] > \beta\,|y_T[\hat{\alpha}(v,.)-b]|\,, \text{since } 1>\beta>0$$

This implies:

$$(1-y_T)[\hat{\alpha}(0,.)-b^*]\frac{\delta\hat{\alpha}(0,b^*,.)}{\delta e} > \beta\left|y_T[\hat{\alpha}(v,b^*,.)-b^*]\frac{\delta\hat{\alpha}(0,b^*,.)}{\delta e}\right| \text{since,} \frac{\delta\hat{\alpha}(.)}{\delta e} > 0$$

$$\Rightarrow (1-y_T)[\hat{\alpha}(0,b^*,.)-b^*]\frac{\delta\hat{\alpha}(0,b^*,.)}{\delta e} > \beta\left|y_T[\hat{\alpha}(v,b^*,.)-b^*]\frac{\delta\hat{\alpha}(v,b^*,.)}{\delta b^e}\right|, \text{since } \frac{\delta^2\hat{\alpha}}{\delta v\delta e} < 0$$

Therefore, it follows that $\frac{\delta g(b^*,.)}{\delta e} > 0$.

Effect of an increase in v:

$$\frac{\delta b^*}{\delta v} = \frac{\frac{\delta g(b^*,.)}{\delta v}}{1 - \frac{\delta g(b^*,.)}{\delta b}} \quad \text{where,}$$

$$\frac{\delta g(b^*,v,.)}{\delta v} = \frac{1}{N}\left[\beta\int_0^{y_T}(\hat{\alpha}(v,b^*,.)-b^*).\frac{\delta\hat{\alpha}(v,b^*,.)}{\delta v}dy\right]$$

Starting from a status quo position of $v = 0$ consider a marginal increase in v targeted to low income households with $y \leq y_T$. The denominator is positive. Consider the numerator.

$$\frac{\delta g(b^*(0),0,.)}{\delta v} = \frac{1}{N}\left[\beta\int_0^{y_T}(\hat{\alpha}(0,.)-b^*(0))\frac{\delta\hat{\alpha}(0,.)}{\delta v}dy\right] = \frac{\beta}{N}y_T(\hat{\alpha}(0,.)-b^*(0))\frac{\delta\hat{\alpha}(0,.)}{\delta v}$$

Since $b^*(0) = \frac{\hat{\alpha}(0,.)}{2}$, $\hat{\alpha}(0) > b^*(0)$. It follows that $\frac{\delta g(b^*,0,.)}{\delta v} < 0$.

Step 2: *Equilibrium cutoff ability at each income level increases with effort and decreases with vouchers.*

Equilibrium cutoff ability level at each income is given by $\hat{\alpha}(y; b^*, e, v, t, c)$.
$\frac{\delta\hat{\alpha}(y;b^*,.)}{\delta e} = \frac{\delta\hat{\alpha}(y;b^*,.)}{\delta e}|b^* + \frac{\delta\hat{\alpha}(y;b^*,.)}{\delta b}.\frac{\delta b^*}{\delta e}$ and $\frac{\delta\hat{\alpha}(y;b^*,.)}{\delta v} = \frac{\delta\hat{\alpha}(y;b^*,.)}{\delta v}|b^* + \frac{\delta\hat{\alpha}(y;b^*,.)}{\delta b}.\frac{\delta b^*}{\delta v}$.
Using step 1 and the signs of the partial derivatives of $\hat{\alpha}$ from 3.1.1, it follows that $\frac{\delta\hat{\alpha}(y;b^*,.)}{\delta e} > 0$ and $\frac{\delta\hat{\alpha}(y;b^*,.)}{\delta v} < 0$.

Step 3: Noting that $N(e, v, b^*, .) = \beta y_T \hat{\alpha}(e, v, b^*, .) + (1-\beta) y_T \hat{\alpha}_0 + (1 - y_T)\hat{\alpha}(e, 0, b^*, .)$, the proof follows from step 2.

Claim 3: *At the household equilibrium under the voucher system, (i) the number of public school students decreases with an increase in β and (ii) the marginal number of students that the public school can gain with an increase in effort increases with β.*

Proof: The proof follows since

$$N_\beta(b^*, .) = -y_T[\hat{\alpha}_0 - \hat{\alpha}(b^*, v, .)] < 0 \text{ and } N_{e\beta}(b^*, .) = y_T \frac{\delta\hat{\alpha}(b^*, v, .)}{\delta e} > 0$$

Proof of Proposition 1. Under the voucher program, e_V solves the first order condition:
$\frac{\delta R_V(e,v)}{\delta e} = (l - c_N(.))N_e(e, v) - C_e(e) = 0$. Since $N_e(.) > 0$ and $C_e(.) > 0$, $l - c_N(.) > 0$ at equilibrium.
Comparative statics with respect to v yields:

$$\frac{\delta e}{\delta v} = \frac{-[(l - c_N)N_{ev} - c_{NN}N_v N_e]}{(l - c_N)N_{ee} - c_{NN}N_e^2 - C_{ee}} \qquad A.1$$

The denominator is negative from the strict concavity of the rent function. Therefore effort increases or decreases under the voucher equilibrium iff $[(l - c_N)N_{ev} - c_{NN}N_v N_e] > 0$. From the strict convexity of the cost function and proposition 2, it follows that $c_{NN}N_v N_e < 0$.

$$N_{ev} = \int_0^1 \left[\frac{\delta^2\hat{\alpha}(y, b^*, .)}{\delta e \delta v} + \frac{\delta^2\hat{\alpha}(y, b^*, .)}{\delta e \delta b}\frac{\delta b^*}{\delta v}\right] dy$$

$\frac{\delta b^*}{\delta v} < 0$ from lemma 1. It can be easily seen that $\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta b} < 0$ and $\frac{\delta^2 \hat{\alpha}(y,b^*,.)}{\delta e \delta v} < 0$. Therefore although the first term is negative, the second is positive and $N_{ev} \gtrless 0$. Therefore A.1 $\gtrless 0$. ■

Proof of Proposition 2.

Proof to part (i):

The rent function under the voucher system is given by:

$$R_V = (p - l)N_0 + lN(e, v, \beta, .) - c_1 - c(N(e, v, \beta, .)) - C(e)$$

The corresponding first order condition is given by:

$$[l - c_N(e, v, \beta, .)]N_e(e, v, \beta, .) - C_e(e) = 0$$

Comparative statics with respect to β yields:

$$\frac{\delta e}{\delta \beta} = -\frac{[l - c_N(.)]N_{e\beta}(.) - c_{NN}(.)N_\beta(.)N_e(.)}{[l - c_N(.)]N_{ee}(.) - c_{NN}(.)N_e^2(.) - C_{ee}(.)}$$

The denominator is negative from strict concavity of the rent function. From claim 3 and the convexity of the cost function it follows that $\frac{\delta e}{\delta \beta} > 0$.

Proof to part (ii):

Using the first order condition in the proof of part (i), comparative statics with respect to l yields:

$$\frac{\delta e}{\delta l} = -\frac{N_e(.)}{[l - c_N(.)]N_{ee}(.) - c_{NN}(.)N_e^2(.) - C_{ee}(.)} \quad \text{which is positive.}$$

It follows that $e_{V,II} > e_{V,I}$.

■

Proof of corollary 1. Suppose phase II is characterized by both a higher β and a higher l, that is $\beta_2 > \beta_1$ and $l_2 > l_1$. Then,

$$\frac{\delta R_{V,II}}{\delta e}|_{e=e_{V,I}} = \frac{\delta R_{V,II}}{\delta e}|_{e=e_{V,I}} - \frac{\delta R_{V,I}}{\delta e}|_{e=e_{V,I}}$$

$$= [l_2 - c_N(N(e_{V,I}, \beta_2, .))]N_e(e_{V,I}, \beta_2, .) - [l_1 - c_N(N(e_{V,I}, \beta_1, .))]N_e(e_{V,I}, \beta_1, .)$$

Since $N_\beta < 0$, $c_N(N(e_{V,I}, \beta_2, .)) < c_N(N(e_{V,I}, \beta_1, .))$. Therefore, $l_2 > l_1$ imply $[l_2 - c_N(N(e_{V,I}, \beta_2, .))] > [l_1 - c_N(N(e_{V,I}, \beta_1, .))]$. Moreover claim 3 implies $N_e(e_{V,I}, \beta_2, .) > N_e(e_{V,I}, \beta_1, .)$. It follows that $e_{V,II} > e_{V,I}$.

Denoting the change in the proportion β, equilibrium quality and equilibrium effort from phase I to phase II, by $d\beta$, dq and de respectively, it follows that $dq = \frac{\delta q}{\delta e}de + \frac{\delta q}{\delta b}[\frac{\delta b}{\delta e}de + \frac{\delta b}{\delta \beta}d\beta]$. The first term in the right hand side is positive and represents the increase in quality in phase II due to an increase in effort. The second term $(\frac{\delta q}{\delta b}(\frac{\delta b}{\delta e}de))$ is positive—an increase in effort increases peer quality which in turn increases quality in phase II in comparison to the first phase. The last term is negative. An increase in the proportion β decreases peer quality which in turn decreases quality. However the last term is small and is more than offset by the first two positive terms, so that quality increases in the second phase. The intuition behind the last term is as follows. Peer quality in the MPS in a voucher regime can be thought of as a weighted average of the average abilities of two groups—the group that choose to remain in the MPS even in spite of vouchers and the group that is forced to return to the MPS because they are unsuccessful in getting a seat at a voucher private school. An increase in β does not affect the average ability of the first group. It also does not affect the average ability of the second group—this is because the students coming back represent a *random* sample of those that want to go and their average ability will always equal the average ability of the group that want to leave with vouchers irrespective of what the proportion β is. However the weightage of the two groups changes and an increase in β decreases the weightage of the second group. Since the second group has a higher average ability, an increase in β decreases the MPS peer quality. In terms of the Milwaukee voucher program, the weight of the second group is substantially smaller than that of the first group and the fall in weightage has been very small (around 0.04). Therefore quality should be expected to increase in the second phase as compared to the first. ■

References

Camilli, Gregory and Bulkley, Katrina (2001), "Critique of 'An Evaluation of the Florida A-Plus Accountability and School Choice Program'," Education Policy Analysis Archives, Vol. 9, Number 7.

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," mimeo, Harvard University.

Chay, Kenneth, Patrick McEwan and Miguel Urquiola (2003), "The central role of noise in evaluating interventions that use test scores to rank schools," NBER Working Paper # 9354.

Epple, Dennis and Richard Romano (1998), "Competition between Private and Public Schools, Vouchers and Peer Group Effects," American Economic Review 62(1), 33-62.

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

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

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

Harris, Doug (2001), "What Caused the Effects of the Florida A+ Program: Ratings or Vouchers?" in Martin Carnoy (ed.), School Vouchers: Examining the Evidence, Economic Policy Institute.

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

Hoxby, Caroline M. (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 (2003), " When Schools Compete, how do they Compete? An assessment of Chile's nationwide School Voucher Program," NBER Working Paper # 10008.

Innerst, Carol (2000), “Competing to win: How Florida’s A-Plan has Triggered Public School Reform,” Urban League of Greater Miami, The Collins Center for Public Policy, Floridians for School Choice, the James Madison Institute and the Center for education Reform.

Kupermintz, Haggai (2001), “The Effects of vouchers on School Improvement: Another Look at the Florida Data”, Education Policy Analysis Archives, Volume 9, Number 8.

Manski, Charles F. (1992), “Educational Choice (Vouchers) and Social Mobility,” Economics of Education Review, 11(4), 531-542.

McMillan, Robert (2004),“Competition, Incentives, and Public School Productivity,” Journal of Public Economics, 88, 1871-1892.

Nechyba, Thomas J. (1996), “Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers,” NBER working paper # 5642.

Nechyba, Thomas J. (1999), “School Finance Induced Migration Patterns: The case of Private School Vouchers,” Journal of Public Economic Theory 1 (1), 5-50.

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

Nechyba, Thomas J. (2003), “Introducing School Choice into Multi-District Public School Systems”, in Caroline Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

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

Wisconsin Legislative Audit Bureau (1995), “An evaluation of Milwaukee Parental Choice Program,” Report 95-3.

Wisconsin Legislative Audit Bureau (2000), “Milwaukee Parental Choice Program,” Report 00-2.

Witte, John F. (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

Year	Number* of Schools	Aid ** Members	MPS Enrollment	Voucher	MPCP Amount (Millions)	Funding of the MPCP Amount: MPS Reduction (Millions)	MPS % of Aid	All other Districts Reduction (Millions)	All other Districts % 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: 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 4: 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 5: Milwaukee Parental Choice Program
Choice Students, by Grade, All Schools (School Year 1994-95)

Grade Level	Number of Students	% of Total Students	Grade Level	Number of Students	% of Total Students
Kindergarten (4-year olds)	73	8.8%	7th	45	5.4%
Kindergarten (5-year olds)	120	14.4	8th	41	4.9
1st	148	17.8	9th	29	3.5
2nd	85	10.2	10th	23	2.8
3rd	81	9.8	11th	8	1.0
4th	71	8.6	12th	9	1.1
5th	53	6.4	Total	830	100
6th	44	5.3			

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

Table 6: Effect of the Milwaukee Program on the 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*prog	-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	0.22	0.16	-0.69***	-1.46	0.43	0.89*
*trend	(0.32)	(0.15)	(0.27)	(0.90)	(1.12)	(0.51)
Somewhat treated*program	0.70	-0.12	-0.89***	-1.21***	-0.02	1.06***
*trend	(0.25)	(0.13)	(0.20)	(0.39)	(0.32)	(0.32)
More treated*program	0.08	-0.39***	0.61***	-0.29	-0.80***	1.38***
*trend	(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 7: Pre-program trend of the different treated groups (as compared to the control group)
WRCT Grade 3 Reading (1989-90), ITBS Grade 5 Math (1987-90), Reading (1987-90) and Language Arts (1989-90)

	WRCT				ITBS					
	% above		% below		Reading		Math		Language Arts	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
trend	-3.84	-4.34**	4.02**	5.80***	-4.09	-3.45	-3.04*	-2.52**	5.44**	4.89*
	(2.33)	(2.16)	(1.79)	(2.08)	(4.11)	(3.42)	(1.66)	(0.98)	(2.71)	(2.66)
More treated * trend	-3.08	-2.03	1.57	0.89	3.01	1.88	0.56	0.32	-2.84	-1.81
	(3.41)	(3.35)	(2.71)	(3.13)	(3.69)	(2.73)	(1.97)	(1.40)	(3.72)	(3.75)
Somewhat treated * trend	-4.41	-3.61	3.84	2.28	3.14	2.12	0.73	0.31	-3.96	-4.29
	(3.01)	(2.67)	(2.34)	(2.44)	(4.05)	(3.17)	(1.83)	(1.21)	(2.48)	(3.71)
Less treated * trend	-2.33	-3.23	-0.29	-1.24						
	(3.61)	(3.10)	(2.53)	(2.71)						
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	242	242	242	242	411	411	410	410	207	207
R-squared	0.50	0.87	0.40	0.83	0.30	0.56	0.30	0.71	0.29	0.83

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. This table uses the 66-47 sample. Huber-White standard errors are in parenthesis. OLS regressions include more treated, somewhat treated and less treated dummies. Fixed effects regressions include school fixed effects. Controls include race, sex, percentage of students eligible for free or reduced price lunch and real per pupil expenditure.

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*program	0.64		-0.26		1.22		0.61		6.28	
*trend	(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. Columns (1), (3), (7), (9) include a time trend, program dummy, program dummy interacted with trend, while columns (2), (4), (6), (8), (10) 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 in Phase I using a continuous treatment variable
(Sample of Milwaukee Public Schools)

	WRCT				ITBS					
	% above		% below		Reading		Math		Lang 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	0.00		0.00		-0.02		0.00		0.00	
* trend	(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 10: 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	0.04*		0.03		-0.09		0.01		-0.13	
* trend	(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 10, 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 11: Effect of Voucher Program on Treatment Status, Milwaukee Phase II

(WKCE Grade 4 Reading, Language Arts, Math, Science and Social Studies, 1997-2002)

	Reading		Language 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 * prog*trend	-0.46		-0.19		0.44		1.24		
	(1.07)		(1.13)		(1.05)		(1.09)		
Somewhat treated* prog	0.34		0.13		1.17		0.92		
* trend	(0.66)		(0.69)		(0.76)		(0.79)		
More treated * prog*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 and 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 12: Comparing the impact of the Milwaukee voucher program in Phase I with that in Phase II

Using performance in reading test [WRCT, ITBS and WKCE] and math test [ITBS and WKCE]

Panel A	Using Treatment Groups							
	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)
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 B	Using Continuous Treatment Variable							
	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)
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 the Panel A sample: Standard deviation of WRCT (% above) scores = 16, Standard deviation of WRCT (% below) Scores = 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 the Panel B sample: Standard deviation of WRCT (% above) scores = 15.81, Standard deviation of WRCT (% below) Scores = 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.

Table 13: Mean Reversion in Wisconsin, 1997-1998.

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

Panel A	Method 1: Based on total score rank									
	Reading		Language Arts		Math		Science		Social Science	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
low*trend	-1.34	-0.26	-3.59	-2.14	2.22	3.87^{*}	1.68	3.08	3.62	5.30^{**}
	(3.47)	(2.58)	(3.09)	(2.02)	(3.65)	(2.17)	(3.24)	(2.09)	(2.95)	(2.13)
mid*trend	-2.37	-1.36	-3.52	-3.14	2.19	3.63^{**}	1.90	1.85	0.75	0.44
	(2.99)	(2.02)	(2.68)	(1.63)	(3.33)	(1.86)	(3.15)	(1.98)	(2.51)	(1.67)
high*trend	-3.74	-2.40	-3.49	-2.80	0.08	0.98	-2.95	-3.79	-2.44	-2.37
	(4.02)	(2.49)	(4.47)	(2.63)	(5.36)	(3.01)	(4.30)	(2.46)	(3.69)	(2.31)
Observations	229	229	230	230	229	229	230	230	230	230
R^2	0.53	0.91	0.58	0.93	0.56	0.92	0.64	0.94	0.67	0.93

Panel B	Method 2: Based on individual subject score rank									
	Reading		Language Arts		Math		Science		Social Science	
	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-reduced lunch 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 14: Comparing the impact of the Milwaukee voucher program in Phase I with that in Phase II, after correcting for mean reversion

Panel A	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)
More Treated * 1 year after prog	-0.06	0.14	0.06	0.20	-0.48**	0.33**	-0.24	0.00
More treated * 2 years after prog	0.38*	-0.38*	0.36	0.50**	0.00	0.65**	0.26	0.11**
More treated * 3 years after prog	0.35	0.03	0.15	0.53**		0.69***	-0.13	0.30***
More treated * 4 years after prog	0.69*	-0.42*		0.50**		0.53***	-0.22	0.25***

Panel B	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)
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***

All figures are in terms of respective sample standard deviations and pertain to the 66-47 sample. For relevant standard deviations, see notes for tables 13 and 14. Mean reversion in panels A and B are respectively based on methods 1 and 2 respectively.

Table A.1: Effect of the Voucher program on treatment status, Phase I and Phase II
Checking robustness using different samples

(WRCT % above scores, 1989-1997 and WKCE Reading Scores 1997-2002)

	WRCT % above			WKCE Reading		
	60-47 FE	66 FE	60 FE	60-47 FE	66 FE	60 FE
	(1)	(2)	(3)	(4)	(5)	(6)
Less treated * 1 year after program	2.55			2.95		
	(3.20)			(2.83)		
Less treated * 2 years after program	2.81			-0.16		
	(2.67)			(2.52)		
Less treated * 3 years after program	3.24			1.28		
	(3.75)			(2.25)		
Somewhat treated * 1 year after program	0.90	2.64	1.56	2.94	2.66	2.89
	(3.02)	(2.63)	(2.75)	(1.93)	(1.79)	(1.84)
Somewhat treated * 2 years after program	4.57**	5.22**	4.24**	4.44**‡	3.29*	3.18
	(2.64)	(2.25)	(2.36)	(2.10)	(1.89)	(1.95)
Somewhat treated * 3 years after program	6.00**	6.42**	5.47**	3.50*	2.99*	2.79
	(3.21)	(2.90)	(3.01)	(2.01)	(1.78)	(1.84)
More treated * 1 year after program	-0.72	-0.90	0.76	2.39	2.72	2.41
	(2.93)	(3.32)	(2.92)	(2.14)	(2.36)	(2.13)
More treated * 2 years after program	6.35**	5.94*	6.32**	5.87***††	6.54***	5.89***
	(2.68)	(3.12)	(2.68)	(2.18)	(2.41)	(2.17)
More treated * 3 years after program	5.52	5.73	5.60	6.33***†††	6.94***+	6.35***+
	(3.47)	(3.97)	(3.46)	(2.22)	(2.54)	(2.22)
Observations	1195	1195	1195	669	669	669
R-squared	0.58	0.57	0.57	0.74	0.79	0.79

*, **, ***: 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.
Huber-White standard errors are in parenthesis. All columns include school fixed effects, year dummies and control for race, sex, % of free-reduced lunch population and real per pupil expenditure.

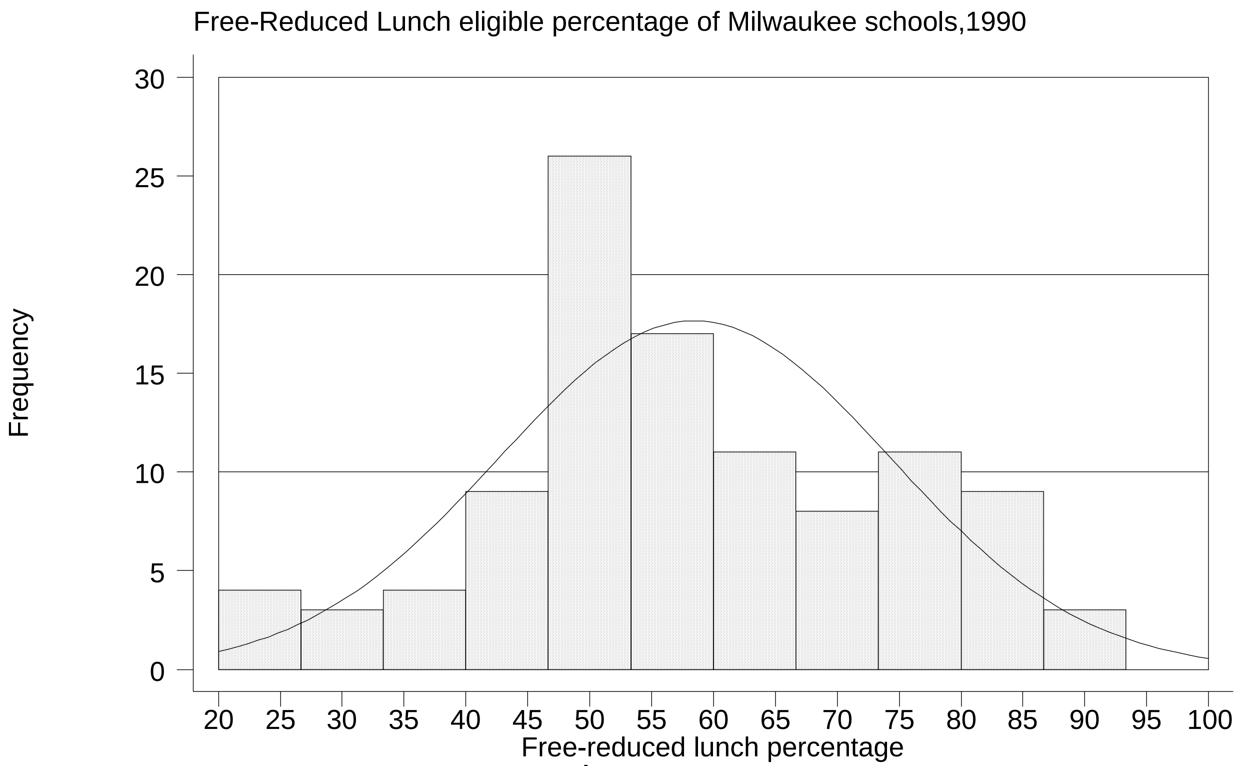
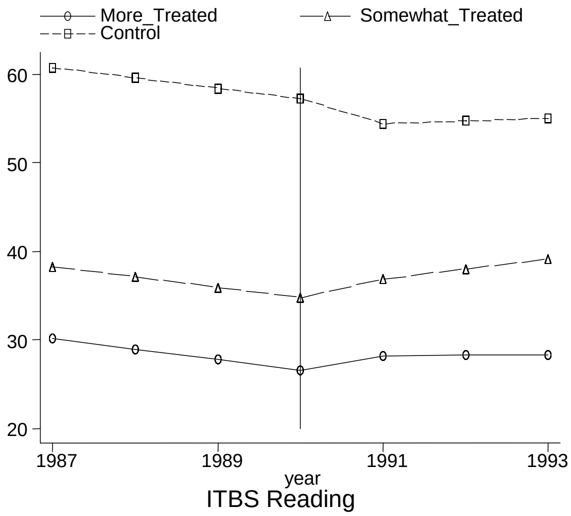
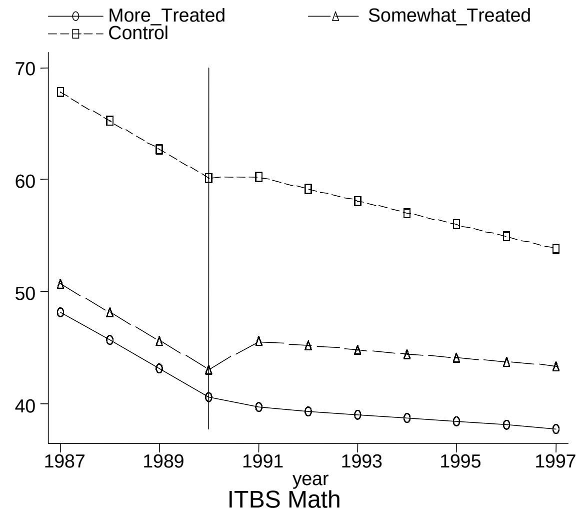

Free-Reduced Lunch eligible percentage of Milwaukee schools,1990
Frequency
30
25
20
15
10
5
0
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Free-reduced lunch percentage

Figure 1

Sample of More treated, Somewhat treated and Control Schools





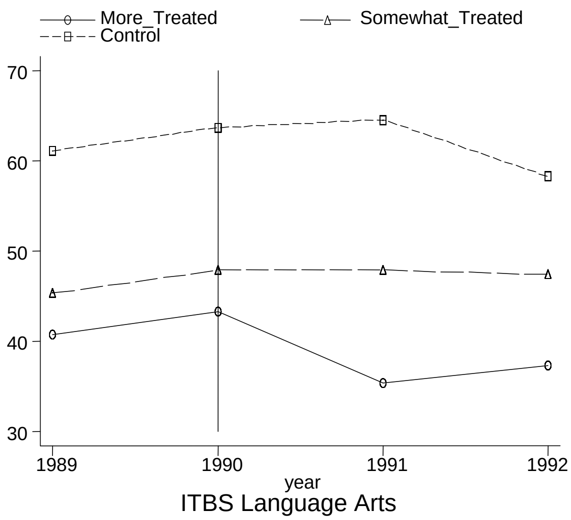


Figure 2. Milwaukee Voucher Program, Phase I

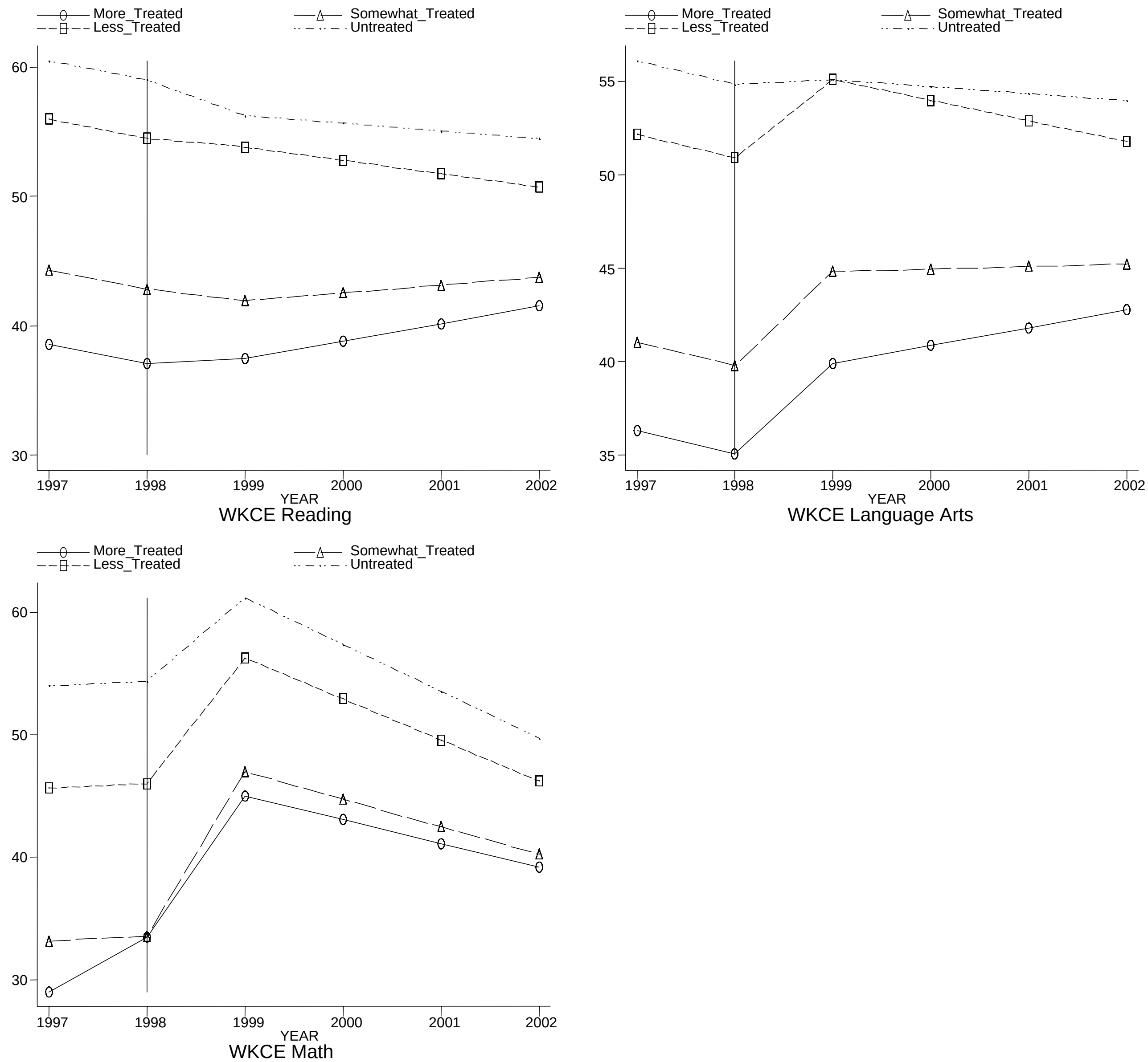


Figure 3. Milwaukee Voucher Program, Phase II