

Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs*

Rajashri Chakrabarti[†]

Cornell University

This version: October 2003

Abstract

This paper examines the impact of vouchers in general and voucher design in particular on public school performance. It argues that all voucher programs are not created equal. There are often fundamental differences in voucher designs that affect public school incentives differently and induce different responses from them. It analyzes two voucher programs in the U.S. The 1990 Milwaukee experiment can be looked upon as a “voucher shock” program that suddenly made low-income students eligible for vouchers. The 1999 Florida program can be looked upon as a “threat of voucher” program, where schools getting an “F” grade for the first time are exposed to the threat of vouchers, but do not face vouchers unless and until they get a second “F” within the next three years. In the context of a formal theoretical model, the study argues that the threatened public schools will unambiguously improve under the Florida-type program and this improvement will exceed that under the Milwaukee-type program. Using school-level scores from Florida and Wisconsin, and a difference-in-differences estimation strategy in trends, it then shows that these predictions are validated empirically.

*I am grateful to Steve Coate, Ron Ehrenberg and Miguel Urquiola for helpful comments, suggestions and encouragement. I thank George Jakubson, Dean Lillard and Mike Waldman for helpful discussions. I would also like to thank the seminar participants at the Labor Economics Workshop, Joint Applied Microeconomics and Public Economics Workshop at Cornell University for useful comments and discussions. All errors are my own.

[†]Department of Economics, Cornell University, Ithaca, New York 14853. Email:rc66@cornell.edu

1 Introduction

The 1983 report “A Nation at Risk”¹ and a series of similar reports have led to continued concern that American public schools may be lagging behind their counterparts in other parts of the developed world. This has led to a wave of demands for public school reform. School choice and accountability in general, and vouchers in particular, are among the most hotly debated instruments of public school reform. This paper is motivated by the need to understand the effect of vouchers and, in particular, the designs of different kinds of vouchers on public school performance. It argues that all voucher programs are not created equal. There are often fundamental differences in voucher designs that affect public school incentives differently and in turn bring about different responses from them.

The last decade has seen three publicly funded voucher experiments in the U.S. The first was initiated in Milwaukee in 1990. This was followed by Cleveland in 1996 and Florida in 1999. Interestingly, there are crucial differences in the designs of these programs. The Milwaukee and Cleveland experiments are similar. (In the rest of the paper, I will concentrate on the Milwaukee experiment.²) These two experiments can be looked upon as “voucher shock” programs with a sudden government announcement that the low-income public school population is eligible for vouchers. In particular, starting in the 1990-91 school year, the Milwaukee Parental Choice Program (MPCP) makes all public school students with family income at or below 175% of the poverty line eligible for vouchers to attend non-sectarian private schools.

On the other hand, the Florida program, also known as the Opportunity Scholarship program, makes all students of a public school eligible for vouchers or “opportunity scholarships” if the school gets two “F”³ grades in a period of four years. Thus the Florida program can be looked upon as a “threat of voucher”, rather than a “voucher shock” program. Here the failing schools are first threatened with vouchers and vouchers are implemented only if they fail to meet a government designated cutoff quality level. In particular, a school getting an “F” for the first time is exposed to the threat of vouchers but does not face vouchers unless and until it gets a second “F” within the next three years.⁴ This paper argues that these differences in voucher

¹ National Commission of Excellence in Education (1983), “A Nation at Risk: The Imperative for Educational Reform,” Washington, DC: U.S. Government Printing Office.

² I will focus on the Milwaukee experiment up to 1996-97. This is because following a 1998 Wisconsin Supreme Court ruling, there were some major changes and the program entered into its second phase in 1998-99. Moreover, comparable test score data are available only up to 1996-97.

³ The Florida Department of Education classifies schools according to five grades: - A, B, C, D, F (A-highest, F-lowest).

⁴ In 1999, 78 schools got an “F”. Students in 2 of those schools became eligible for vouchers. No school became eligible for vouchers in 2000 or 2001 although 4 elementary schools got an F in 2000. In 2002, 64 schools got an F. Students in 10 of those schools became eligible for vouchers.

designs will affect public school incentives differently and will induce very different responses from them. In particular, it argues that the Florida type “threat of voucher” program will induce a much greater public school response than the Milwaukee type “voucher shock” program.

Apart from the above differences, the designs of the two programs are strikingly similar. In both the experiments, the private schools are not permitted, by law, to discriminate between students who apply with vouchers—they have to accept all students unless oversubscribed and have to pick students randomly once they are oversubscribed. The system of funding is also very similar. Under each program the average voucher amount equals the state aid per pupil, and the vouchers are financed by an equivalent reduction of state aid to the district. Thus state funding is directly tied to student enrollment and enrollment losses due to vouchers are reflected in a revenue loss for the public school. The average voucher amounts under the Florida (1999-2000 through 2001-2002) and Milwaukee (1990-1991 through 1996-1997) programs have been respectively \$3,330 and \$3,346. The corresponding revenue loss per student as a percentage of total revenue per pupil has been 41.66% in Florida and 45.23% in Milwaukee during this period.

The paper develops its argument in the context of a formal theoretical model with three agents: – the public school, which aims at maximizing net revenue (which I call “rent”); the households, which care about school quality; and the private schools, which are represented by exogenously given quality levels. The demand for public school is endogenously determined from household behavior, giving micro-foundations to the public school payoff function. In an equilibrium framework, the model endogenously determines public school quality and its ingredients – public school effort and peer group quality. Both under complete information and under moral hazard (when public school effort is not observable), the model generates two empirically testable predictions that hold at the respective program equilibria – the threatened public schools will show an unambiguous improvement in quality under the Florida-type “threat of voucher” program and the improvement under the “threat of voucher” program will exceed that under the Milwaukee-type “voucher shock” program.

Using school-level test score data from Florida and Wisconsin, the paper next proceeds to test the two theoretical predictions. Implementing a difference-in-differences estimation strategy in trends, it estimates the program effects for each of the experiments by comparing the post-program improvement of the treated schools 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 both the

theoretical predictions. These findings are quite robust in that they continue to hold after controlling for other confounding factors such as mean reversion. Several strategies, including a regression discontinuity estimation strategy, are used to address the potential problem of mean reversion.

Although the main motivation for writing this paper is to highlight how differences in voucher designs can have radically different effects on public school incentives and performance, there is another motivation that is no less important. It is very difficult to empirically quantify the pure public school response to vouchers or even to determine its direction. On the one hand, vouchers affect public school quality directly through public school response; on the other, vouchers lead to sorting which affects public school quality through changes in student composition and the public school peer group.⁵ All these three factors are reflected in the observable public school outcome - school scores. Therefore an empirical analysis of school scores captures a combination of all these factors rather than the pure public school effect. As a result, we would ideally want an experiment where the only response to vouchers is from the public school and there is no sorting, and consequently no effect on student composition or peer group. The Florida program, to a large extent, provides such a case. In Florida, the schools threatened by vouchers respond by optimizing on effort, but there is no sorting if the school manages to avoid another “F”. Therefore, an empirical analysis of the improvement of these schools promises to provide an idea about the direction and magnitude of the pure public school response.

The main theoretical studies relevant to this paper are Epple and Romano (1998) and McMillan (2001). Epple and Romano look at the effect of vouchers on the choice between public and private schools. They argue that vouchers lead to sorting by income and ability. However, they take public schools as passive and thus abstract from the incentive effects of vouchers on public school response. McMillan models public school behavior. He shows that under certain circumstances, public schools may find it optimal to reduce productivity when a voucher is introduced. There are some important differences between his paper and this study. The main difference is that he considers the effect of traditional voucher experiments (“voucher shock” in my terminology) on public school response. On the other hand, this paper compares and contrasts the effects of two types of voucher experiments – a “voucher shock” and a “threat of voucher” on public school performance. Second, this paper derives the demand for public school from equilibrium household behavior, thus providing micro-foundations to the public school payoff function, unlike McMillan. Third,

⁵ See Hsieh and Urquiola (2003)

since households differ only by income in McMillan’s model, he is unable to model peer-group quality.⁶ In this paper, households are characterized by an income-ability tuple. Peer-group quality is endogenously determined which enables me to examine the effect of vouchers on peer group quality and explicitly model the interactions of peer-group quality, household choices, and public school choice.

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”). There are three such studies in the context of Milwaukee. Witte et al. (1995) find no evidence of any effect; Greene et al. (1996) find positive effects in both reading and math, while Rouse (1998) finds positive effect in math, but no effect in reading for Milwaukee choice students. Surprisingly, only two studies thus far look at the effect of vouchers on public school performance in the U.S.. Greene (2001) looks at the effect of the Florida program while Hoxby (2001) looks at the effect of the Milwaukee program after the Supreme Court ruling of 1998. Both studies find evidence of a positive productivity response to vouchers. However, both studies use only two years of data – one pre-program and one post-program year – and hence cannot control for any pre-program trend. Thus, public school response to vouchers is a relatively neglected area both in the theoretical and empirical literature. In particular, there is no study thus far (either theoretical or empirical) that seeks to compare the public school response to different voucher designs. This study fills this important gap.

2 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,⁷ which I call “rent” in rest of the paper. Rent is simply defined as revenue minus costs. Public school revenue is given by $p \cdot N$,⁸ where p is the exogenously given per pupil

⁶ Since peer group quality is an important ingredient of school quality, understanding the effect on peer group quality is essential to understanding the effect on school quality.

⁷ I adhere to the general line of thought in the school competition literature [Hoxby (2001), McMillan (2001)] that the public school maximizes net revenue. 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.

⁸ This formulation captures the fact that revenue is directly tied to the number of students under each of the programs as well as in the simple public-private (baseline) system. However, as discussed earlier, in both Florida and Milwaukee, the public school loses only the state aid per pupil for each student lost due to vouchers. Therefore, a more appropriate formulation would be to model revenue as a more general function of enrollment $p(N)$. For simplicity, I assume a multiplicative form. All results go through with the more general functional form $p(N)$, $p'(N) > 0$.

revenue and N is the number of students in public school. Public school cost (C_p) depends on its effort (e) and number of students, $C_p(N, e) = c(N) + C(e)$. Both $c(\cdot)$ and $C(\cdot)$ functions are assumed to be increasing and strictly convex in their respective arguments. I assume $p - c_N > 0$, that is the “net marginal revenue” per student is positive.

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.⁹ Households pay a tuition $T = t \cdot Q$ ($t > 0$) to attend a private school of quality Q .¹⁰

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 $U(x, \theta, \alpha) = h(x) + \alpha u(\theta)$. The functions h and u are increasing and strictly concave in x and θ respectively. It follows that households with higher ability have a higher preference (marginal valuation) for school quality, $U_{\theta\alpha} > 0$.¹¹

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 $e \in [e_{min}, e_{max}]$ ¹² and public school peer quality b . Public school peer quality is defined as the mean ability of the public school student body.¹³ If a public school household decides to switch to a private school with vouchers, it incurs a positive switching or relocation cost c .

The paper models three alternative scenarios: (i) a simple public-private system (PP) without vouchers (the baseline), which can be thought of as the pre-program scenario for both programs; (ii) the Milwaukee-type “voucher shock” (VS) program; and (iii) the Florida-type “threat of voucher” (TOV) program. The simple public-private system consists of two stages. In the first stage, the public school chooses effort. In

⁹ 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.)

¹⁰ Note that at equilibrium, private school quality will always exceed public school quality. Otherwise, no household would pay to attend a private school.

¹¹ The assumption $U_{\alpha\alpha} = 0$ is made for simplicity. All results go through under $U_{\alpha\alpha} < 0$, $U_{\theta\alpha\alpha} < 0$ (and thrice differentiability).

¹² Public School effort is assumed to be observable here for simplicity. As appendices B and C show, the results continue to hold when effort is unobservable, but what is observable is a noisy representation of effort (for example, public school scores).

¹³ 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.

stage 2, households choose between schools after observing the last stage public school effort.¹⁴ 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 . In stage 2, facing v , the public school chooses effort. In stage 3, households choose between schools (after observing v and e) and incur switching costs if they transfer out of public school. Peer-group quality and public school quality are simultaneously obtained.

The Florida program is modeled in four stages. In the first stage, the Government announces the program and a corresponding cutoff quality \bar{q} and voucher v . In stage 2, facing the program the public school chooses effort. Given the existing peer group quality,¹⁵ quality q is realized. In stage 3, the government imposes vouchers v if $q < \bar{q}$. No voucher is imposed if $q \geq \bar{q}$. In the last stage, households choose between schools (after observing effort and whether vouchers were imposed) and incur switching costs if they transfer out of public schools. Peer-group quality and public school quality are simultaneously realized.

Each of the systems constitutes a game between two players: the public school and the households. The TOV program is initiated by a government announcement that all students of a school will be eligible for vouchers v if it fails to meet a certain quality cutoff \bar{q} .¹⁶ Facing this program, and knowing its existing peer group quality, the public school deduces the effort level \bar{e} that it needs to match to escape vouchers. Facing this \bar{e} cutoff, and correctly anticipating household behavior, it chooses effort to maximize rent. Given the existing peer group quality, quality is realized simultaneously. In stage 3, the government imposes vouchers or not, as promised, based on the observed quality. In stage 4, based on observed public school effort and whether vouchers were imposed in stage 3, the households anticipate a certain peer-group quality and choose between schools. At equilibrium, anticipated peer quality equals actual peer quality. This yields an equilibrium peer quality and a corresponding allocation of households between public and private sectors. Equilibrium public school quality (which is a composite of equilibrium public school effort and peer quality) is simultaneously obtained. Therefore an equilibrium of the “threat of voucher” program is an effort-peer quality tuple (e_{TOV}, b_{TOV}) , such that given the quality cutoff \bar{q} and voucher v (i) e_{TOV} is a public school equilibrium, given b_{TOV} and (ii) b_{TOV} is a household equilibrium, given e_{TOV} .

¹⁴ The general notion in all the three systems is that households observe last year’s scores and whether vouchers were given and then choose between schools.

¹⁵ Since the public-private system is the pre-program scenario, the existing peer-group quality is the public-private peer group quality.

¹⁶ Note that since peer group quality and effort are observable, this is equivalent to announcing a cutoff in terms of effort.

Under the Milwaukee-type program and the simple public-private system, the behavior of the households and public school are similar. Households observe previous stage public school effort and v ($v = 0$ in the public-private system) and decide whether or not to move. The public school moving before the households correctly anticipates household behavior and makes its rent maximizing effort choice. 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 shock” equilibrium is a peer-group quality b_{VS} and an effort e_{VS} such that given voucher v (i) e_{VS} characterizes the public school equilibrium, given b_{VS} and (ii) b_{VS} characterizes the household equilibrium, given e_{VS} .

3 Characterization of the program equilibria

This section solves for the household and public school equilibria and compares the public school qualities under the PP, VS, and TOV equilibria. Subsection 3.1 considers household behavior while subsection 3.2 considers public school behavior.

3.1 Household behavior

This subsection analyzes the household behavior under the three systems in a common framework. Under each of the three systems, households observe public school effort and whether vouchers were imposed and make their utility maximizing decisions. Each household can either choose to go to the public or to a private school. In the former case, it gets utility $h(y) + \alpha u(q(e, b))$. In the latter case it gets utility $h(y + v - t \cdot Q^* - c) + \alpha u(Q^*)$,¹⁷ where Q^* is the optimal private school quality choice of household (y, α) given v, t and c . A household (y, α) chooses private school iff $h(y + v - t \cdot Q^* - c) + \alpha u(Q^*) > h(y) + \alpha u(q(e, b))$. Define $D = [h(y + v - t \cdot Q^* - c) + \alpha u(Q^*)] - [h(y) + \alpha u(q(e, b))]$. It can be easily seen that $\frac{\delta D}{\delta y} > 0$ and $\frac{\delta D}{\delta \alpha} > 0$ which imply stratification by income and ability respectively. Since these also hold under $v = 0$, even a simple public-private system (without vouchers) is characterized by stratification by income and ability.

Suppose all households expect a peer group quality $b^e \in [0, 1]$. I assume that there are always some households in the public and some households in the private sector for each income level.¹⁸ Then for each

¹⁷ The parameter v takes on a value of zero under the pre-program public-private system, and under the Florida TOV system if the public school escapes vouchers. On the other hand, v takes on an exogenously given positive value under the VS program, and under the TOV program if the public school fails to meet the cutoff and vouchers are introduced.

¹⁸ This assumption is made for simplicity. All results go through as long as there is at least one income for which this

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:

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

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 $D_\alpha > 0$, by the implicit function theorem,

$$\hat{\alpha} = \hat{\alpha}(y; v, e, b^e, t, c) \quad (3.1.1a)$$

is a continuously differentiable function.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 .²¹ Given all other parameters, the cutoff ability level varies inversely with y . This is because both higher income and higher ability increase preference towards private schooling. Given b^e , peer group quality b is given by:

$$b = \frac{\int_0^1 \int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha dy}{\int_0^1 \int_0^{\hat{\alpha}(y, b^e, \cdot)} d\alpha dy} \Rightarrow \frac{1}{2} \cdot \frac{\int_0^1 \hat{\alpha}^2(y, b^e, \cdot) dy}{\int_0^1 \hat{\alpha}(y, b^e, \cdot) dy} = g(b^e, e, v, t, c) \quad (3.1.2)$$

At equilibrium b corroborates the initial conjecture b^e , that is,

$$b = b^e \quad (3.1.3)$$

In other words, if all households expect a peer-group quality, then at equilibrium this expectation has to be fulfilled. Mathematically, given parameters e, v, t, c , a fixed point in b is reached.

Proposition 1 *A household equilibrium always exists.*

All proofs are in appendix A. From (3.1.1)-(3.1.3), the equilibrium peer quality satisfies the equation $b^* = g(b^*, e, v, t, c)$. The corresponding equilibrium allocation of households between public and private sectors 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 Similarly, for each α and given t, v, e, c, b^e there exists a unique household \hat{y} such that all households with lower income choose public school and those with higher income choose private school. Again using implicit function theorem $\hat{y} = \hat{y}(\alpha; v, e, b^e, t, c)$ is a continuously differentiable function and the cutoff income level (for each ability) is decreasing in v and increasing in e, t, c, b^e .

²¹ $\frac{\delta \hat{\alpha}}{\delta v} = \frac{-h_x^Q}{u(Q^*) - u(q)} < 0$, $\frac{\delta \hat{\alpha}}{\delta e} = \frac{u_x^q}{u(Q^*) - u(q)} > 0$, $\frac{\delta \hat{\alpha}}{\delta b^e} = \frac{u_{b^e}^q}{u(Q^*) - u(q)} > 0$, $\frac{\delta \hat{\alpha}}{\delta t} = \frac{Q^* \cdot h_x^Q}{u(Q^*) - u(q)} > 0$, $\frac{\delta \hat{\alpha}}{\delta c} = \frac{h_x^Q}{u(Q^*) - u(q)} > 0$, $\frac{\delta \hat{\alpha}}{\delta y} = \frac{h_x^q - h_x^Q}{u(Q^*) - u(q)} < 0$. Superscript Q or q refers to the value of the partial derivative when the household chooses private or public school respectively.

is characterized by $\hat{\alpha}(y, b^*, \cdot)$ for $y \in [0, 1]$. $N(b^*, e, v, t, c) = \int_0^1 \int_0^{\hat{\alpha}(y, b^*, \cdot)} d\alpha dy = \int_0^1 \hat{\alpha}(y, b^*, \cdot) dy$ gives the corresponding number of students in public school at the household equilibrium b^* . The rest of this section will explore how changes in the different parameters affect equilibrium peer quality, allocation of households between public and private sectors, and number of public school students. It should, however, be first pointed out that while a household equilibrium always exists, it may not be unique. To see this differentiate (3.1.2) with respect to b^e to find that

$$\frac{\delta g}{\delta b^e} = \frac{1}{N(b^e, \cdot)} \cdot \int_0^1 (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}}{\delta b^e} dy \text{ where } N(b^e, \cdot) = \int_0^1 \hat{\alpha}(y, b^e, \cdot) dy$$

The sign of $\frac{\delta g}{\delta b^e}$ is positive.²² If it exceeds one, there are multiple equilibria. 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 where $b(0) > 0$ and $\frac{\delta g}{\delta b^e} < 1$. These are sufficient conditions that ensure a unique equilibrium. The first condition always holds since $0 < \hat{\alpha}(\cdot) < 1$. The second condition implies that a small increase (decrease) in anticipated peer quality leads to a less than proportionate increase (decrease) in actual peer quality. It holds 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. The intuition behind this is simple. If ($\frac{\delta u}{\delta q}$) and ($\frac{\delta q}{\delta b^e}$) are large, a small increase in b^e leads to a large increase in utility from the public school. This attracts a substantial number of high-ability households to the public school (that is, it leads to a large outward shift in $\hat{\alpha}(y; \cdot)$) so that peer quality increases more than in proportion to anticipated peer quality and there are multiple equilibria.

Lemma 1 *Equilibrium peer group quality falls with vouchers and increases with public school effort.*²³

An increase in public school effort leads to an influx of higher ability households at each income level, thus leading to an increase in equilibrium peer quality. On the other hand, vouchers lead to a flight of higher ability households at each income level, thus decreasing peer quality.

Lemma 2 *Vouchers increase sorting by income and ability. An increase in public school effort leads to reverse sorting by both income and ability.*

²² The proof is in appendix A.

²³ A similar analysis shows that equilibrium peer quality increases with increases in t and c .

An increase in public school effort leads to an increase in the equilibrium cutoff ability level, $\hat{\alpha}(y, b^*)$, at each income level. This occurs through two channels. Given b^* , 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. Vouchers acting directly as well as indirectly through peer quality induce a flight of high ability public school households at each income level to the private sector at equilibrium (sorting by ability). Symmetrically, an increase in effort leads to an influx and vouchers a flight of high income households at each ability level.²⁴

Proposition 2 *Equilibrium number of public school students decreases with vouchers and increases with public school effort.*

Noting that $N(b^*, e, v, t, c) = \int_0^1 \hat{\alpha}(y; b^*, \cdot) dy$ the proof follows directly from the proof of Lemma 2.²⁵ Since an increase in effort (vouchers) leads to an influx (flight) of households at each income level, equilibrium number of students²⁶ increases (decreases) with effort (vouchers). Note that lemmas 1, 2 and proposition 2 apply to each of the three systems under (PP, VS, TOV) consideration.

3.2 Public School Behavior

The public school correctly anticipates behavior in all the future stages of the corresponding game, and chooses effort to maximize rent. The rent function²⁷ is given by $pN(e, v) - c(N(e, v)) - C(e)$.²⁸ Under the PP system there exists a unique effort e_{PP} such that it solves the first order condition $\frac{\delta R(e, 0)}{\delta e} =$

²⁴ An interesting point to note is that the driving force behind sorting here is different from that in Epple and Romano (1998). In Epple and Romano, sorting arises due to an interplay of both demand and supply side factors. Private schools attract high ability students with tuition discounts while high-income households prefer private schools to public. Here, in keeping with the feature of the Milwaukee-Cleveland-Florida experiments, private schools are not allowed to discriminate between students. Rather, they are modeled as passive. Therefore sorting here arises exclusively due to the demand side factor that higher income and higher ability households have a higher preference towards school quality. A related point to note here is that due to the absence of the supply side factor (the discriminatory private schools), sorting or “cream skimming” is likely to be less in these voucher experiments than the traditional voucher experiments (example, Chile) where private schools can choose between students.

²⁵ It can be easily seen that the cutoff ability level (at each income), the cutoff income level (at each ability) and the equilibrium number of public school students increases with increases in t and c .

²⁶ The analysis here assumes that when vouchers are imposed, all households, irrespective of income, become eligible for them. Although this is the case in Florida, in Milwaukee vouchers are targeted only to the low-income population. I abstract from this here for simplicity. As appendix B of Chakrabarti (2003a) shows, all results continue to hold under targeted vouchers. Note that given other parameters (e, v, t, c) , the number of students is less in a household equilibrium where all households are eligible rather than where only the low-income are. The obvious reason is that in the former case there is a flight of households at each income level, whereas in the latter case it is restricted only to a subset of income levels.

²⁷ 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. Since the cost function is strictly convex, the rent function is strictly concave under this assumption.

²⁸ Note that N depends on other parameters t, c also but they are suppressed to simplify notation.

$(p - c_N)N_e(e, 0) - C_e(e) = 0$. Similarly under the VS program, there exists a unique effort e_{VS} such that it solves the first order condition $\frac{\delta R(e, v)}{\delta e} = (p - c_N)N_e(e, v) - C_e(e) = 0$.

Lemma 3 *Vouchers decrease rent.*

This lemma is applicable to both the VS and the TOV systems. It says that given other parameter values, an imposition of vouchers decreases rent. This is because vouchers lead to a loss of students to the public school. Given that the net marginal revenue is positive, this leads to a decrease in rent.

Proposition 3 *Equilibrium public school effort under the “voucher shock” 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: $[(p - c_N)N_{ev} - c_{NN}N_v N_e]$ (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 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 consists of two effects. $N_{ev} = \int_0^1 \left[\frac{\delta^2 \hat{\alpha}(y, b^*, \cdot)}{\delta e \delta v} + \frac{\delta^2 \hat{\alpha}(y, b^*, \cdot)}{\delta e \delta b} \frac{\delta b^*}{\delta v} \right] dy$. 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.²⁹ 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 ($\frac{\delta^2 \hat{\alpha}(y, b^*, \cdot)}{\delta e \delta b} < 0$). 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). This implies that the marginal number of students and hence net marginal revenue can either increase or decrease with vouchers, thus rendering the effect on public school effort ambiguous. Public school effort increases if either net marginal revenue increases or the decrease in marginal revenue is less than the

²⁹ 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.

decrease in marginal cost. In other words, it increases iff the new marginal revenue exceeds the new marginal cost.

Proposition 4 *For each voucher v , there exists a cutoff effort level \bar{e} ³⁰ such that the equilibrium effort³¹ under the “threat of voucher” program, e_{TOV} exceeds both*

- (i) *the equilibrium effort under the “voucher shock” program, e_{VS} and*
- (ii) *the equilibrium effort under the public-private system, e_{PP} .*

The Florida-type TOV program affects public school incentives in a way very different from the Milwaukee-type VS program. A Florida public school facing the threat has two options: it can choose to meet the cutoff or it can choose not to meet the cutoff. In the latter case, it is in the same state as its counterpart under the VS program. It chooses the VS optimum effort e_{VS} and gets the VS rent $R(e_{VS}, v)$. Since vouchers decrease rent, it follows that the school can be induced to satisfy a cutoff \bar{e} strictly higher than e_{VS} , where the rent from \bar{e} without vouchers exactly equals the rent from e_{VS} with vouchers. Thus, the fundamental feature of the TOV that induces a higher effort is that vouchers are not already imposed and a sufficient improvement can enable schools to escape vouchers.³² Note that any cutoff in the range $(e_{VS}, \bar{e}]$ induces an effort under the TOV program that is strictly higher than under the VS program. The intuition behind the second part of the proposition is similar. The Florida TOV program introduces a discontinuity in the rent function at the cutoff effort level. If the cutoff is set at e_{VS} , then meeting it gives a higher rent than choosing to accept vouchers. Since e_{PP} is the rent maximizing effort under $v = 0$, setting the cutoff at e_{PP} gives an even higher rent to the public school. Given the strict concavity of the rent function, this implies

³⁰ Note that since peer quality is known, announcing a cutoff in terms of effort is equivalent to announcing a corresponding cutoff in terms of quality.

³¹ As appendices B and C show, these results continue to hold when effort is not observable. But, as is obvious, the cutoff can no longer be set in terms of effort, which is now unobservable.

³² The analysis here assumes that all households irrespective of income are eligible for vouchers under the VS program. However, this result also holds for vouchers targeted to the low-income population under the VS (Chakrabarti 2003a). The intuition can be laid down in two steps. Call the VS program where all students are eligible the “universal voucher shock” (UVS) program and where only the low-income students are eligible the “targeted voucher shock” (TVS) program and the corresponding equilibrium number of students and equilibrium effort N_{UVS} , N_{TVS} and e_{UVS} , e_{TVS} respectively. First, note that the equilibrium rent under the TVS is greater than that under the UVS. Under the TVS, the school can attract N_{UVS} students by giving a lower effort than under the UVS (follows from footnote 26 and proposition 2) and hence at a rent higher than under the UVS. Since the school chooses to attract N_{TVS} , it must be the case that rent is higher under the TVS. Second, if vouchers when imposed in the Florida-type TOV program took a targeted form, then following the argument in proposition 4, the program could implement a cutoff $\bar{e} > e_{TVS}$. But vouchers take the universal form in Florida, which implies that the rent would be smaller than the TVS rent if schools failed to meet the cutoff. This implies that there exists a cutoff $\bar{e} > \bar{e} > e_{TVS}$ which satisfies the school’s incentive constraint with equality and hence can be implemented by the TOV program. To summarize, there are two features in the design of TOV that induce a higher effort than the TVS: (i) vouchers are not already imposed and (ii) the potential loss of students is much greater.

that there exists a cutoff $\bar{e} > e_{PP}$ which satisfies the school’s incentive constraint. Again, any cutoff in the range $(e^*, \bar{e}]$ induces an effort under the TOV program that is strictly higher than under the PP program.

In the TOV program it may be reasonable to think that there is a stigma attached to being labeled as a ‘voucher public school’. For example, Maureen Backentoss, assistant superintendent of curriculum and instruction of Lake County School District refers to it as a “glass of cold water in the face”. In the presence of such a stigma, the public schools gain an additional utility if they are able to escape vouchers. This feature is absent in the VS program. Note that this will weigh results in favor of the TOV and will induce an even higher improvement under the TOV.³³

Corollary 1 *(i) Equilibrium public school quality under the “threat of voucher” equilibrium.³⁴*

(a) exceeds the equilibrium quality under the pre-program public-private system.

(b) exceeds the equilibrium quality under the “voucher shock” program.

(ii) Equilibrium public school quality under the “voucher shock” program can be greater or less than the pre-program public-private equilibrium quality.

From proposition 4, the effort under the TOV program exceeds that under the VS and PP equilibria. Noting that peer quality varies positively with effort and vouchers tend to lower peer quality, part (i) of the proposition follows. Since equilibrium effort under the VS program can be greater or less than the pre-program public-private equilibrium, the same follows for equilibrium quality.³⁵

4 Empirical Strategy

The empirical part of the paper seeks to test the following two predictions obtained from the theoretical model:

(i) A threatened public school in the Florida-type “threat of voucher” program will respond by increasing effort, which will result in an increase in its quality.

(ii) Quality improvement of a threatened public school in the Florida-type program will exceed the improvement (if any) of a treated public school in the Milwaukee-type “voucher shock” program.

³³ Another interesting property of the TOV program that is not a part of the VS program is that a mere ‘threat’ of vouchers (rather than vouchers) can induce an increase in public school effort.

³⁴ An implicit assumption in rest of the paper will be that the Government appropriately designs the program, that is \bar{e} lies within the relevant effort ranges.

³⁵ Note that if equilibrium effort falls with vouchers, quality will be unambiguously less under the VS equilibrium compared to the PP equilibrium.

Under the 1999 Florida opportunity scholarship program, if a school gets two failing grades in a period of four years, all its students become eligible for vouchers. The schools that received an “F” grade in 1999 were directly exposed to the threat of vouchers. These schools will constitute the group of treated schools and will be referred to as the “F schools” from now on. The schools which got a grade of “D” in 1999 were closest to the F schools in terms of grade and at the same time were not directly treated by the program. These schools will constitute the group of control schools and will be referred to as the “D schools”. The treatment and control groups respectively consist of 65 and 457 elementary³⁶ schools. Since the program was announced in June 1999 and the grades were based on the tests held in February 1999, the classification of schools into treatment and control groups is made here on the basis of their pre-program scores and grades.

This paper considers public school scores as the outcome variable. The identifying assumption here is that if the F schools and D schools have similar trends in the pre-program period, any shift of the F schools compared to the D schools 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-program data.

$$s_{it} = f_i + \beta_0 t + \beta_1 (F * t) + \gamma X_{it} + \epsilon_{it}$$

where s_{it} is i th school score in year t , f_i are school fixed effects, t denotes time trend, F is a dummy variable taking a value of 1 for F schools and 0 for D schools, $F * t$ is an interaction between F dummy and trend, X_{it} denotes the set of school characteristics and ϵ_{it} is a stochastic error term. β_1 captures the pre-program difference in trend of the F schools.³⁷

If the F schools and D schools have the same pre-program trend, I use the following set of specifications to investigate whether the F schools demonstrate a higher improvement in test scores compared to the control D schools in the post-program era. If the treated F schools demonstrate a differential pre-program trend, in addition to estimating these specifications, I also estimate a slightly modified version of them where I control for their pre-program differences in trends. I begin with a completely linear model:

$$s_{it} = f_i + \alpha_0 t + \alpha_1 v + \alpha_2 (F * v) + \alpha_3 (v * t) + \alpha_4 (F * v * t) + \alpha_5 X_{it} + \epsilon_{it} \quad (1)$$

³⁶ I shall restrict my analysis to the elementary schools as there were too few middle and high schools that received a grade of “F” in 1999 (7 and 5 respectively) to justify analysis.

³⁷ When there are data on more than two pre-program years (as in the case of Florida writing) I also fit a non-linear specification with pre-program year dummies and interactions of F dummy with pre-program year dummies. This allows the individual F school pre-program year effects to vary in an unrestricted way from those of the D schools.

where v is the program dummy, $v = 1$ if year > 1999 and 0 otherwise. The coefficient α_0 denotes the common pre-program trend. 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 $F * v$ and $F * v * t$ estimate the program effects – α_2 captures the intercept shift and α_4 the trend shift of F schools compared to the D schools after the program.³⁸ All specifications I describe here are fixed effects regressions. I also estimate OLS counterparts of each of these specifications.

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.

$$s_{it} = f_i + \beta_{99}D_{99} + \beta_{00}D_{00} + \beta_{01}D_{01} + \beta_{02}D_{02} + \beta_0(F * v) + \beta_1(F * v * t) + \beta_2X_{it} + \epsilon_{it} \quad (2)$$

where D_{99} , D_{00} , D_{01} and D_{02} are year dummies for 1999, 2000, 2001, and 2002 respectively. β_0 and β_1 reflect the post-program intercept and trend shifts respectively of the F schools 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 F school year dummy to capture individual post-program year effects.

$$s_{it} = f_i + \gamma_{99}D_{99} + \gamma_{00}D_{00} + \gamma_{01}D_{01} + \gamma_{02}D_{02} + \gamma_0(F * D_{00}) + \gamma_1(F * D_{01}) + \gamma_2(F * D_{02}) + \gamma_3X_{it} + \epsilon_{it} \quad (3)$$

This specification no longer constrains the post-program year-to-year gains of the F schools to be equal and allows the program effect to vary across the different years. The coefficients γ_0 , γ_1 , γ_2 represent the effect of one, two and three years into the program respectively for the F schools.

The above specifications assume that the D schools are not affected by the program. The D schools do not face any direct threat from the program. However, since the D schools are close to getting “F”, they may face an indirect threat and hence may be induced to improve. In fact, there is some anecdotal evidence that D schools may have responded to the program. None of the schools in the Hillsborough County School District received an F rating in 1999, although 37 schools received a D rating. Hillsborough’s superintendent

³⁸ The coefficients α_2 and α_4 represent the program effects under the assumption that D schools do not respond to the program. If the D schools also respond, α_2 and α_4 will be underestimates of the program effect.

announced that he would take a 5% pay cut if any of his schools received an “F” on the next school report card.³⁹ Similarly, Martin County School District, which had no F school in 1999, announced \$1,000 bonus for teachers at D schools who meet new performance goals. Therefore, I allow the F and D schools to be different treated groups (with varying intensities of treatment) and compare their post-program improvements, if any, with a suitable control group. The schools which received a “C” grade from the state in 1999, being the next higher up in the grade scale, constitute my control group here. They will be referred to as the “C schools” from now on. I estimate the following modified versions of the above specifications considering F and D schools as two treated groups (more treated and less treated respectively) and C schools as the control group.

$$s_{it} = f_i + \alpha_0 t + \alpha_1 v + \alpha_2 (F * v) + \alpha_3 (D * v) + \alpha_4 (v * t) + \alpha_5 (F * v * t) + \alpha_6 (D * v * t) + \alpha_7 X_{it} + \epsilon_{it} \quad (1')$$

$$s_{it} = f_i + \sum_{i=1999}^{2002} \beta_i D_i + \beta_0 (F * v) + \beta_1 (F * v * t) + \beta_2 (D * v) + \beta_3 (D * v * t) + \beta_4 X_{it} + \epsilon_{it} \quad (2')$$

$$s_{it} = f_i + \sum_{i=1999}^{2002} \gamma_{0,i} D_i + \sum_{i=2000}^{2002} \gamma_{1,i} (F * D_i) + \sum_{i=2000}^{2002} \gamma_{2,i} (D * D_i) + \gamma_3 X_{it} + \epsilon_{it} \quad (3')$$

where D_i represents year dummies.

I next consider the Milwaukee program. The extent of treatment of the Milwaukee public schools depends on the percentage of their students eligible for free or reduced price lunches.⁴⁰ Therefore, I classify the schools into different treatment groups based on their pre-program (1989-90 school year) percentage of free or reduced-price lunch eligible students. Figure 1 shows the distribution of Milwaukee elementary⁴¹ public schools according to the percentage 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% were black and 14.81% were hispanic. In the somewhat treated group an average of 53.6% were

³⁹ See Innerst (2000).

⁴⁰ 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. Since the cutoff of 175 % is not strictly enforced and households within this 10% margin are often allowed to apply, the extent of treatment is proxied here by the percentage of free or reduced-price lunch eligible students of the Milwaukee public schools.

⁴¹ The analysis here is restricted to elementary schools only because there are very few middle and high schools in the MPS.

free or reduced-price lunch eligible, 50.57% were black and 36.8% 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 mean⁴² and some schools above and below the mean, I construct a second sample, the “66-47” sample. 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. 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.⁴³

For each of these samples, the control group consists of Wisconsin schools outside Milwaukee which satisfied the following criteria in 1990: (i) were urban, (ii) had at least 25% of their population eligible for free or reduced-price lunch, and (iii) black students constituted at least 15% of the school population⁴⁴. Since the Wisconsin schools outside Milwaukee are much more advantaged than the Milwaukee schools, I could find only 33 elementary schools which satisfied these characteristics in 1990. 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. Using each of these samples, I investigate how the different treatment groups in Milwaukee responded to the VS program. For this purpose, I first test the underlying identifying assumption that the pre-program trends of the untreated and the different treatment groups are the same. Second, I estimate OLS and fixed-effects versions of the three specifications (1)-(3) after adjusting for the number of treatment groups.

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. There is a wide variation among Milwaukee schools in the percentage of their

⁴² The mean percentage of free or reduced-price lunch students in the Milwaukee Public Schools in 1990 was 59%.

⁴³ 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.

⁴⁴ The control group criteria used here are based on Hoxby (2001).

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 investigate whether an increase in the intensity of treatment is associated with higher improvement (in terms of intercept, trend, and individual-year shifts) after the program. In this approach the program effect is assumed to vary linearly with treatment intensity. To relax this assumption, I also estimate models that contain interactions between treatment intensity and dummy variables representing different treatment groups.

5 Data

The data for this paper come from multiple sources. The Florida data consist of school-level data on test scores, socio-economic characteristics of schools, and school finance data and are obtained from the Florida Department of Education (DOE). Socio-economic characteristics include data on ethnicity and sex (1994-2002)⁴⁵, percentage of students eligible for free or reduced-price lunches(1997-2002) and are obtained from the school indicators database of the Florida DOE. School finance data consist of several measures of school level and district level per pupil expenditures and are obtained from the school indicators database and the Office of Funding and Financial reporting, Florida DOE.

School-level data on test scores are available on two tests: (i) the Florida Comprehensive Assessment Test Sunshine State Standards (FCAT-SSS) (This test will be referred to as the FCAT in the remainder of the paper.) (ii) the Stanford 9 test which the state calls the FCAT Norm Referenced test (FCAT-NRT). Following a field test in 1997,⁴⁶ the FCAT reading and math tests were administered in the year 1998 in grades 4, 8, and 10 for reading and grades 5, 8, and 10 for math. Mean scale scores (on a scale of 100-500) are available for reading (grades 4, 8, and 10) and math (grades 5, 8, and 10) for 1998-2002. The Florida writing test (known as the Florida Writes test before 2000) was first administered in spring 1993 in grades 4, 8, and 10. Mean scale scores (on a scale of 1-6) are available for writing in grades 4, 8 and 10 for 1994-2002. The Stanford 9 test is administered in all grades 3-10 in reading and math only. This test compares the performance of Florida students with those of students across the nation. This test was first administered in 2000, so I do not have any pre-program score for this test. School-level data on mean scale scores (on a

⁴⁵ For the remainder of this paper, I will refer to school years by the calendar year of the spring semester.

⁴⁶ The purpose of the 1997 field test was to evaluate the quality of the test items before they were included on a test on which students received scores.

scale of 424-863) and NPR (national percentile rank) scores are available for all grades 3-10 in reading and math for 2000-02. (The FCAT is a high-stakes test, unlike the Stanford 9, because only the scores from the former enter the calculation of school grades.)

The Wisconsin data consist of school-level data on test scores, socio-economic characteristics of schools, and per pupil expenditure (both school-level and district-level data). The data 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 ethnicity, sex, and percentage of students eligible for free or reduced-price lunches for the period 1989-97 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. 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] and (ii) the Iowa Test of Basic Skills (ITBS). The WRCT is a state-administered grade 3 reading test that has been administered since 1989⁴⁷. School scores are reported in three “performance standard categories” for this test: percentage of students below, percentage of students at, and percentage of students above the standard.⁴⁸ WRCT data for these three categories are available for 1989-97. The ITBS reading and math tests were administered by the district during the period under consideration. Starting with 1994, the ITBS was administered only in math; as of 1999, the ITBS was no longer administered as a district assessment program. School-level ITBS reading data are available for 1987-1993; ITBS math data are available for 1987-1997.⁴⁹

6 Results

The results are arranged in the following order in this section. Subsections 6.1 and 6.2 analyze the program effects in Florida and Milwaukee respectively. Subsection 6.3 compares the two program effects. Subsection 6.4 discusses the issue of mean-reversion. The last subsection reports other robustness checks.

⁴⁷ The mode of reporting WRCT scores changed in 1998. The test score data prior to 1998 are not comparable to those for 1998 and after. So I focus on scores up to 1997.

⁴⁸ Percentage of students below the performance standard, percentage of students at the standard, and percentage of students above the standard will be denoted by % below, % at, and % above, respectively, in the remainder of the paper.

⁴⁹ The mode of reporting ITBS math scores changed in 1998. So I focus on pre-1998 scores.

6.1 Florida

Table 1 compares the pre-program trend of the treated F schools with that of the control D 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. All regressions are weighted by the number of students tested; present estimates with Huber-White standard errors; and include ethnicity, sex, percentage of students eligible for free or reduced-price lunches, and real per pupil expenditure as controls. Although there is no significant evidence of any difference in pre-program trend between the two groups in reading and math, F schools exhibit a significant negative differential trend in writing.

Table 2 presents the effect of the Florida TOV program on F school reading scores as compared to the D schools. Columns (1) and (2) fit the completely linear model 1. After controlling for post-program common intercept and trend shifts, these columns show evidence of a positive and significant intercept shift of F schools after the program. There is also a positive trend shift, but it is no longer significant in the fixed-effects estimate. Columns (3) and (4) allow the comparison group trend to be non-linear and present results from model 2. The results are similar and confirm the effects in the first two columns. These effects are disaggregated in columns (5) and (6) where the coefficients reflect the effects of the program after one, two and three years (model 3). Both the OLS and fixed-effects estimates show positive and significant year effects in each of the years after implementation.

Table 3 looks at the effect on math and writing scores. Columns (1) and (2) estimate model 1 and show a positive, significant, large intercept shift after the program although there is no evidence of any trend shift. The results from model 2 in the next two columns corroborate this evidence. Columns (5) and (6) (model 3) show evidence of positive significant F school year effects in each of the three years after the program. In writing, columns (7) through (10) show evidence of a large, positive and significant intercept shift, and a positive, significant trend shift for the F schools. The last two columns show positive, significant year effects in each of the three years after the program.

To summarize, using different specifications and all three subjects tested (reading, math, and writing), tables 2 and 3 underline a robust finding in Florida: the treated F schools show considerable improvement after the program as compared to the D schools. Figure 2 graphs the OLS estimates from the linear model. It confirms that 1999 has been the watershed year. The F schools improved relative to the D schools after

the program was implemented, and the gap between F schools and D schools has undoubtedly narrowed.

Since the D schools are close to getting an “F” , they may also face an indirect threat and hence may be induced to improve. In that case the estimates above will be underestimates. To investigate whether D schools have also responded to the program, I consider D schools as an additional treatment group. Designating F schools as the “more treated” group, D schools as the “less treated”, and C schools as the control group, I repeat the above analysis. Tables 4 and 5 present the effects of the program on F and D schools as compared to the C schools in reading, math, and writing.⁵⁰ For each subject, the first two columns present results from model 1', the last two columns from model 3', and the middle column(s) from model 2'. The first two columns in table 4 show a positive significant trend shift for both F schools and D schools. Both treated groups also show a positive intercept shift although it is not significant for the D schools in the fixed-effects estimate. Also both the F school trend and intercept shifts are considerably larger in magnitude than those of D schools. The results seem to be robust in that they are reflected in columns (3) and (4) also. Columns (5) and (6) show positive significant year shifts for both the treated groups in each of the three years after the program. The F school shifts are larger and significantly different from those of the D schools.

In table 5, the first two columns show a positive, significant intercept shift for F schools in math which is statistically greater than that of the D schools. Estimation of model 2 in column (3) confirms this result. Columns (4) and (5) show positive significant year effects for both F schools and D schools. However the F school shifts are significantly higher than those of D schools. In writing, columns (6)-(8) show positive intercept shifts for both F schools and D schools where the F school shifts are significantly different from those of D schools. Although there is no evidence of any trend shift for F schools, it is statistically higher than the negative (though insignificant) trend shift of the D schools. In columns (9)-(10) the F schools show positive year effects in each of the three years after the program. These shifts are significantly different from the corresponding D school shifts.⁵¹ Again, consistent with the finding in the previous columns, the D

⁵⁰ F schools exhibit a negative differential pre-program trend in reading and writing (and no differential trend in math) as compared to the C schools. D schools exhibit a negative pre-program trend in reading and positive pre-program trend in math and writing in comparison to the C schools (Chakrabarti, May 2003). Whenever there is a difference in pre-program trend, tables (4) and (5) control for these differences by including interactions between trend and more treated and less treated dummies respectively. When there are only two years before program (for example, reading and math), the pre-program difference between treatment and control groups can be either a trend difference or a year effect. In reading and math, specifications 1 and 2 control for this pre-program difference assuming it is a trend difference, and specification 3 controls for it assuming this difference is a year effect. Results from regressions without controlling for these pre-program differences are qualitatively similar.

⁵¹ In 2002, although the state still continued to grade schools on the scale of “A” through “F”, the grading criteria was

schools only exhibit a positive significant year effect in the first year after the program.⁵²

To summarize, the results presented in tables 2-5 seem to be quite robust. Using different samples, different subjects, different specifications, and both OLS and FE estimates of each of the specifications, the results reflect the same consistent set of findings:

(i) F schools have responded to the program. In each of the three subjects tested - reading, math, and writing - the treated F schools show considerable improvement after the program in comparison to the control schools.

(ii) The D schools also seem to have responded to the program, which is consistent with the anecdotal evidence. They show non-negligible improvement (in comparison to the C schools) - at least in reading and writing.

(iii) The improvement of F schools has been considerably larger and also statistically different from those of D schools. These findings validate the first prediction obtained from theory.

These results also shed light on the second question that has motivated this paper. In traditional voucher experiments, it is very difficult to separately identify the pure public school response to voucher incentives. This is because the changes reflect three simultaneous phenomenon: public school response, change in composition of students, and change in peer quality. The Florida program provides a unique perspective to this issue. The F schools face the threat of vouchers and hence respond to incentives. However, if they succeed in escaping vouchers then there is no sorting or change in peer composition due to vouchers. None of the 65 F schools got a second “F” during the period 2000-01 and hence satisfy this condition. Therefore an examination of the F school response to the program promises to give us an idea about the pure public school response to vouchers. The foregoing analysis reveals that we should expect the public schools to respond positively to voucher incentives and the magnitude of this response is quite high. Using C schools as controls and given an in-sample standard deviation of 20 in reading, 20 in math and 0.53 in writing, this suggests effect sizes of 0.47, 0.50 and 0.80 in reading, 0.45, 0.55 and 0.60 in math and 0.74, 0.70 and 0.64 in writing after first, second and third years after the program, respectively.⁵³

changed to include value added scores in addition to levels. However, since the grades were still based on FCAT (SSS) scores and the F schools anticipated vouchers if they got a second F in 2002, similar incentives continued to play in 2002. Moreover, the results are very similar if the year 2002 is dropped and the analysis is repeated with data through 2001 only.

⁵² It should be noted here that the F schools and D schools (especially, the F schools) received additional funds from the state. However, all the results above are obtained after controlling for real per pupil expenditure. The results do not change after controlling for a polynomial in real per pupil expenditure.

⁵³ However, it should be noted that these effects might be overestimates of pure public school response to traditional voucher experiments. This is because sufficient improvement in Florida can enable schools to avoid loss of students altogether. Hence

Consistent with the above findings, there is considerable anecdotal evidence that suggest that F schools have responded to the program. I summarize some of them below. Escambia county implemented a 210-day extended school year in its F schools (typical duration was 180 days), implemented an extended school day at least twice a week, added small group tutoring in afternoons and Saturdays and longer time blocks for writing and math instruction.⁵⁴ To curb absenteeism it started an automated phone system to contact parents when a child was absent. Miami Dade County hired 210 additional teachers for its 26 F schools, switched to phonics, and encouraged parents (many of whom were dropouts) to go back to school for a high school equivalency diploma. Broward County reduced its class size to 18-20 students in its low performing schools and beefed up services for children whose main language was not English. Palm Beach County targeted its fourth grade teachers for coaching and began more frequent and closer observations of teachers in its F schools. In the words of Carmen Varela-Russo, associate superintendent of technology, strategic planning and accountability, Broward County Public Schools, “People get lulled into complacency”. . . “the possibility of losing children to private schools or other districts was a strong message to the whole community.” (Innerst, 2000).

6.2 Milwaukee

Using the 66-47 sample, table 6 looks at the effect of the Milwaukee “voucher shock” program on WRCT (% above), WRCT (% below), ITBS reading, and ITBS math scores of different treatment groups.⁵⁵ For each set the first column reports results from model 1, the second column from model 2, and the third column from model 3. All regressions reported in this table include school fixed effects⁵⁶ and control for ethnicity, sex, percentage of students eligible for free or reduced-price lunches, and real per pupil expenditure. Column (1) finds that after controlling for common post-program trend and intercept shifts there is no statistically significant evidence of any trend shift after the program. Although the intercept shifts are positive and non-negligible in magnitude, they are not statistically significant. Moreover the intercept shifts are smallest for the more treated group, which is contrary to what we would expect if they were program effects. Estimates of model 2 in column (2) corroborate these findings (except that the more treated intercept shift now exceeds

the Florida program is likely to provide higher incentives than traditional voucher experiments, which are always associated with some loss of students irrespective of the amount of effort.

⁵⁴ Innerst (2000).

⁵⁵ The pre-program trends of the different treatment groups are not statistically different from those of the control schools (Chakrabarti, 2003a). Therefore, any post-program shift of the treated schools can be attributed to the program.

⁵⁶ The results from OLS estimation of the corresponding models are very similar and hence are not reported here.

the less treated shift although it is still smaller than the corresponding somewhat treated coefficient). These effects are disaggregated in column 3 where the coefficients reflect the effect of the program after one, two and three years after the program for the different treated groups. Although most of the first and third year effects are positive, they are not significant and do not always have the right hierarchy between groups. The second year effect, on the other hand, is positive significant (at least for the more treated and somewhat treated groups) and quite high in magnitude. Interestingly, the more treated effect exceeds the somewhat treated effect, which in turn exceeds the less treated effect (although the difference between the different group effects are never statistically different from zero). The results for WRCT (% below) are qualitatively similar. The second year effect is again statistically significant for both the more treated and somewhat treated groups (although the more treated effect is marginally smaller than the somewhat treated effect.)

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.⁵⁷ As a result, my comparison group here will be 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. Columns (7)-(9) show no evidence of any statistically significant effect of the program on ITBS reading scores of any of the treated groups. However, the effects are always positive, though the more treated effects are not always greater than the somewhat treated ones. The picture for ITBS math is broadly similar. However now the second year effect of the somewhat treated group in the fixed-effects estimate is positive, significant, and exceeds the more treated effect.⁵⁸ An interesting thing to note here is that, similar to the findings for WRCT, the second year effects are the largest and quite high in magnitude for both ITBS reading and math, although they do not have the right hierarchy and are often not statistically different from zero. Note that none of the more treated effects in this table are ever statistically different from the somewhat treated or less treated effects and none of the somewhat treated effects from the less treated effects. Figure 3 graphs the OLS estimates for ITBS from the linear model. As expected there is no evidence of any program effect (MT denotes more treated, ST somewhat treated and LT less treated).

Table 7 considers a continuous treatment variable and proxies the intensity of treatment by the pre-program (1990) percentage of free-reduced lunch students of schools. Columns (1) and (2) control for

⁵⁷ 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.

⁵⁸ Note that although the more treated school effects are jointly significant for the WRCT (% above) scores in models 1 and 2 at 10% level, they are no longer jointly significant either in the non-linear model or for the WRCT (% below) or the ITBS reading and math scores.

common post-program trend and intercept shifts and fit a completely linear model. There is no evidence of any improvement with an increase in treatment intensity. Columns (3) and (4) include year dummies to control for common post-program shocks. The results are very similar to those in the first two columns. Columns (5) and (6) disaggregate these effects into individual year effects. Although the OLS estimate of the second year effect is positive significant, it is no longer so in the fixed-effects estimate and also drops to less than half the magnitude. The analysis in this table assumes that the treatment effect varies linearly with treatment intensity. To relax this assumption, I also include interactions of treatment intensity with different treatment groups (results are not reported here as they do not add any new insight), but none of these coefficients are found to be significantly different from zero.

The findings in tables 6 and 7 can be summarized as follows: The results are mixed. However, it is safe to say that there is no evidence of any negative effect of the “voucher shock” program. Although most of the coefficients are positive they do not always have the right hierarchy and are often not significantly different from zero. However, the program seems to have had a positive and significant effect in the second year after program, at least in WRCT. These results seem to be robust in that they are replicated in the analysis with other samples as illustrated in appendix table B.3. Note that these findings are consistent with the prediction of the theoretical model (proposition 2).

6.3 Florida Versus Milwaukee

Using the results from subsections 6.1 and 6.2, this subsection compares the effect of the Florida and Milwaukee programs on the respective more treated schools. Table B.1 shows that the demographic characteristics of the more treated and control groups in Florida are very similar to those of the corresponding groups in Milwaukee (for both the 66-47 and 60-47 samples).⁵⁹ Table 8 compares the effect of the Florida program with that of the Milwaukee program using reading and math scores. The estimates here are based on those in Tables 4-6 and all figures are expressed in terms of the respective sample standard deviations. The estimates from the linear model 1 (columns 1-3) show positive, significant intercept and trend shifts for the Florida more treated schools and much smaller statistically insignificant effects in Milwaukee. Columns (4)-(6) report results from the non-linear model. The Florida program has positive significant year effects in all the three years after program. In Milwaukee, although most of the year effects are positive, only the

⁵⁹ For the rest of this section I shall focus on the 66-47 sample in Milwaukee. The results for the other samples are qualitatively similar.

second year WRCT effect is statistically significant— and each of the Florida effects exceed the corresponding Milwaukee effects in both WRCT and ITBS. The picture in math in columns (7)-(10) is broadly similar: the Milwaukee effects are never statistically significant and are always less than the corresponding statistically significant effects in Florida. Moreover using the estimates in Appendix table B.3 it can be easily seen that these findings hold for each of the other samples 60-47, 66, and 60 and also for the WRCT (% below scores).⁶⁰To summarize, it can be said that the improvement of the threatened schools in Florida program has been considerably larger than those in Milwaukee.⁶¹ This finding is quite robust since it holds for all the different samples, different specifications, different subjects and different tests. These findings validate the second prediction obtained from theory.

The Milwaukee program saw a major shift in the year 1998-99 and entered into its second phase when following a Wisconsin Supreme Court ruling, the religious schools were allowed to accept choice students for the first time.⁶² Chakrabarti (2003b) shows that, as a result, the improvement of the treated schools in Milwaukee Phase II was much higher than that in Phase I. At first blush, it appears that it would be instructive to check whether the Florida effects exceed those in Phase II also. Unfortunately, the Florida program is not directly comparable to the Milwaukee Phase II program. Due to some funding changes the revenue loss per student is much higher in Phase II than in Florida; the voucher amount is larger; and in Milwaukee we are many more years into the program (9-12 years) since its inception than in Florida (1-3 years). Since each of these factors are expected to increase the Milwaukee effect, it is not clear that the Florida effect will still exceed the phase II effect. Nevertheless, Chakrabarti (2003a) compares the effects of the Florida and the Milwaukee Phase II programs on the respective more treated groups. Interestingly, the findings are similar to those above. For each of the samples, each of the specifications and each of the subject areas, the improvement of the Florida threatened schools exceeds that of the more treated schools in Milwaukee Phase II. This result strengthens the earlier findings and garners further support in favor of theoretical prediction 2.

⁶⁰ The standard deviation for the WRCT(% below) scores is 10.98.

⁶¹ Since many of the coefficients in Milwaukee are not significant though positive, I also do a pair-wise non parametric test (sign test), 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 Florida exceeds the corresponding one in Milwaukee is $\frac{1}{2}$. Since there are three post-program years, I have 9 pair-wise comparisons. Under the null, $D = \text{Florida effect} - \text{Milwaukee effect}$ follows a binomial (9, 0.5) distribution. D is positive in all 9 cases. The probability of getting 9 positive D under the null is 0.002. Since this is very small, the null of equal effects can be comfortably rejected.

⁶² This led to more than a three fold increase in the number of private schools participating in the program and almost a four fold increase in the number of choice students in 1998-99.

6.4 Mean-Reversion

Mean-reversion is the statistical tendency where high or low scoring schools tend to score closer to the mean subsequently. Since the F schools were low scoring in 1999, a natural question to ask would be whether the improvement in Florida is driven by mean reversion rather than the program. Since I do a difference-in-differences analysis, my estimates above will be contaminated by mean reversion only if F schools mean revert to a greater extent than the D schools and/or the C schools.

For a first pass at the mean-reversion issue, table B.2 investigates whether the schools that were low scoring in 1998 were also low scoring in 1999. If there were “good” and “bad” schools, which could be observed with certainty (the “certainty case”), 10% (25%) of the schools will always rank in the bottom 10th percentile (25th percentile) while if schools were equal and were chosen by lottery (the “lottery” case) only 1% (6.3%) of the schools will rank in the lowest 10% (25%) in both years. The actual percentage of schools in the last three columns in panels A and B resemble more closely the certainty numbers than the lottery numbers for each of reading, math, and writing. This implies that although there may be mean-reversion, it may not be a major problem.

A more direct way to approach mean-reversion is to check by how much the schools that received an “F” grade in 1998 improved during the period 1998-1999 compared to those that received a “D” (or “C”) grade in 1998. Since there was no program during this period, this gain can be taken to approximate the mean reversion effect and can be subtracted from the post-program gain of F schools compared to D schools (or C schools) to get at the mean-reversion corrected program effect.⁶³ The accountability system of assigning letter grades to schools started in the year 1999.⁶⁴ However, using the state grading criteria and data on percentage of students in different achievement levels in each of FCAT reading, math and writing,⁶⁵ I was able to assign letter grades to schools in 1998. The schools which were assigned grades “F”, “D” or “C” in 1998 using this method will henceforth be called the 98F schools, 98D schools, and 98C schools, respectively.

Using data for 1998 and 1999, Panel A of table 9 finds that in comparison to the 98D schools, the 98F

⁶³ This is based on the assumption that the mean reversion of F schools compared to D schools (C schools) of that year would be similar for 1999 and 2000.

⁶⁴ The pre-1999 accountability system classified schools into four groups I-IV (I-low, IV-high).

⁶⁵ In FCAT reading and math, the state categorizes students into five achievement levels that correspond to specific ranges on the raw-score scale. Using current year data, the state designates a school an “F” if it is below the minimum criteria in reading, math and writing, a “D” if it is below the minimum criteria in reading or math or writing, and a “C” if it is above the minimum criteria in reading, math and writing (but below the higher performing criteria in all three). In reading and math at least 60% (50%) of the students have to score level 2 (3) and above while in writing at least 50% (67%) have to score 3 and above to meet the minimum (higher performing) criteria in that respective subject.

schools show no evidence of mean reversion either in reading or math although there is mean reversion in writing. In comparison to the 98C schools (Panel B), there is no evidence of any mean reversion in reading; both 98D schools and 98F schools show comparable amounts of mean reversion in math; and only 98F schools show mean reversion in writing.

Table 10 presents the mean-reversion corrected effect sizes of the more treated schools for the first, second and third years after program. These are obtained by subtracting the effect size attributed to mean reversion (obtained from expressing the relevant coefficients in table 9, panel B in terms of respective standard deviations) from the F school effect sizes in each of the three years after program.⁶⁶ The estimates in reading are the same as earlier. In math, although the effect sizes fall, they are still positive and considerably larger than those in Milwaukee.⁶⁷ These results strengthen my earlier findings and further confirms the empirical validity of the theoretical predictions (i) and (ii).

Regression Discontinuity Analysis

The whole issue of mean reversion arises here from the concern that the F schools may revert towards the mean to a greater extent than D schools and/or C schools just by virtue of their relatively low performance in 1999. Therefore, an alternative way to get around the problem of mean reversion is to do a regression discontinuity analysis and compare the improvement of F schools just below the cutoff between “F” and “D” with D schools just above the cutoff. The Florida program has created a highly non-linear and discontinuous relationship between school achievement as embodied in 1999 school grade and the probability that the school’s students become eligible for vouchers in the near future. Treatment depends in a deterministic way on observable 1999 school grade. There exists a sharp cutoff – all F schools face a direct threat of vouchers while the D schools do not face any such direct threat.

To construct discontinuity sample 1, I use the following method to pick F schools and D schools close to the cutoff. First, I rank the F schools based on the sum of the percentage of students below minimum criteria (level 1) in reading, math, and writing⁶⁸. Second, I rank the D schools. The D schools which fail the minimum criteria in maximum number of subjects are considered closest to the F schools⁶⁹ and D schools

⁶⁶ Since none of the F schools got an “F” in either 2000 or 2001, the effect sizes attributed to the Florida program in the second and third years in table 10 may be underestimates.

⁶⁷ A mean-reversion corrected estimate of Florida versus Milwaukee Phase II also gives similar results (Chakrabarti, 2003a).

⁶⁸ I choose this method because the school grades are based on this criteria and ranking by this method will, at least to some extent, resemble/mimic a ranking in terms of grade. The idea is to try to rank the schools in terms of grade.

⁶⁹ I also use an alternative method where I consider all the D schools as a group and rank them in terms of the sum of percentage of students below minimum criteria in all three subjects. The results are similar to those from the above sample.

which fail in only one subject the farthest. Within each subgroup of D schools, I rank the D schools based on the sum of percentage of students below minimum criteria in all three subjects⁷⁰. Third, I pick the F schools and D schools close to the cutoff. The F schools thus chosen fail the minimum criteria in all three subjects while the D schools fail the minimum criteria in at least two and hence the two groups are quite close to each other. The summary characteristics of the F schools and D schools in this sample are shown in Table 11. The F schools and D schools in the discontinuity sample are strikingly similar to each other both in terms of pre-program demographic characteristics and scores and are much closer than the original sample of all F schools and D schools. Using discontinuity sample 1, and specifications (1) and (3), table 12A–Panel A shows that the effect of treatment on reading, math and writing scores are very similar to those obtained earlier. In fact, in some cases the effects exceed those obtained earlier with the sample of all F and D schools. Interestingly, the results appear to be comparable or larger in reading and math and smaller in writing. This is consistent with the earlier finding of mean reversion in writing unlike in reading and math. Next, I progressively shrink the sample of F schools and D schools and using the above ranks I pick F and D schools closer and closer to the cutoff. Panel B shows the results for the linear model for two such samples. The results are qualitatively similar.

To check the robustness of the results I use several other methods to pick F schools and D schools close to the cutoff. First, I construct a discontinuity sample (discontinuity sample 4) where both F schools and D schools fail to meet the minimum criteria in reading and math while in writing only F schools fail the minimum criteria. Here the probability of treatment varies discontinuously as a function of the percentage of students scoring above 3 in FCAT writing. There is a sharp cutoff at 50%. All schools in this sample below 50% face a direct threat, while those above 50% face no such direct threat (see footnote 65). The results from this sample are very similar to those above and are presented in table 12B. I also consider two corresponding discontinuity samples where both groups fail the minimum criteria in reading and writing (math and writing). F schools fail the minimum criteria in math (reading) also, unlike D schools. In these samples, the probability of treatment changes discontinuously as a function of the percentage of students above level 1 in math (reading) and there is a sharp cutoff at 60%. The results are similar and hence are not reported here.

Second, I consider subject specific regression discontinuity samples. For example to estimate the effect

⁷⁰ All classifications here are based on pre-program (1999) all curriculum school scores. Results are qualitatively similar if they are based on standard curriculum school scores.

on reading I consider F schools and D schools such that both failed the minimum criteria in reading. The underlying argument is that although the F schools and D schools lie on opposite sides of the cutoff they are very similar in terms of their pre-program reading scores – any divergence in their post-program reading scores can then be attributed to the program rather than mean-reversion. I consider two other corresponding samples for math and writing respectively.⁷¹ For each of these samples and for each subject and specification the results are similar to those reported above. These imply that the above findings are reasonably robust.

6.5 Some other issues: testing robustness of the results

Is there a stigma effect of getting the lowest performing grade?

The F schools got the lowest performing grade in 1999. If there is a stigma effect of getting the lowest performing grade, the F schools will try to improve merely to avoid this stigma rather than in response to the program. Although the system of assigning letter grades to schools started in 1999, Florida had an accountability system in the pre-1999 period. Starting in 1997, schools were categorized into four groups 1-4 (1-low, 4-high). Using FCAT writing data⁷² for two years (1997 and 1998), I investigate whether the schools, which were categorized in group 1 in 1997, improved in relation to the 1997 group 2 and group 3 schools during the period 1997-98. The rationale here is that if there is a stigma effect of getting the lowest performing grade, the group 1 schools should improve in comparison to the group 2 and 3 schools even in the absence of the TOV program. Table 13 shows that there is no evidence that this has been the case.

Has there been “teaching to the test” in Florida?

The above analysis focuses on high stakes test scores in Florida. Since the threat in the Florida program is given in terms of grade, the response of the Florida threatened schools should be assessed in terms of the high-stakes test. In other words, the theoretical predictions in Florida pertain to performance in the high stakes test; the model has no direct prediction relating to the low stakes test scores or “teaching to the test”. The notion here is that if the public school is found to unambiguously respond to the TOV program by increasing effort, then the other issues of teaching to the test, manipulation of the composition of the test taking population can be more easily taken care of—for example, by broadening the curriculum to include all desirable areas and topics and using the test scores of all grades and students (for example, all special

⁷¹ Note that the regression discontinuity analysis is likely to produce underestimates since the D schools are treated to some extent, although indirectly.

⁷² During this period the local districts used different norm referenced tests for reading and math which varied in content and norms and hence not at all comparable across districts. So I do not use these data for my analysis.

education and limited english proficiency students also) for school grade computation purposes. Moreover, as Hanushek (2002) argues, “teaching to the test” can only have a one-time effect on school scores.

Nevertheless, I investigate this issue by looking at the reading and math scores from the low stakes Stanford 9 test, which was not used in the assignment of school grades. While the Stanford 9 test also contains multiple-choice questions, it places more emphasis on critical analysis in reading and problem-solving strategies, evaluating expressions, and solving linear equations in math compared to the FCAT. Table 14 uses data on Stanford 9 test scores for 2000-02. Since this test was first administered in 2000, no pre-program data are available. Prior to 2000, the districts used a variety of nationally normed tests which not only varied in content but the norms were also different.⁷³ As a result, these data are not comparable across districts, years, or with the post-program Stanford 9 data. Therefore, the pre-2000 data are not used here.

Table 14 panel A shows very high correlation between FCAT and Stanford 9 for both level and change scores for each of the subject areas tested and for each of the F schools, D schools, C schools, and all schools. The implication is that the FCAT results should be replicated in Stanford 9 also. Using Stanford 9 scores for 2000-02, panel B shows that F schools and D schools show positive and significant improvement in all grades and subjects and in most cases the F school effect exceeds the corresponding D school effect, even though the effects are not always statistically different between groups. Note that the effects are likely to be underestimates.⁷⁴ The overall picture is consistent with the FCAT picture earlier. Up to 2001, the FCAT reading and math scores only in grades 4 and 5, respectively, were used for school grade computation. Interestingly, the 2001 F school improvement in reading is largest in grade 4. However for math, the 2001 F school gain is smallest in grade 5. To summarize, from the limited data that are available, the results are mixed. There is some evidence in favor of “teaching to the test”, but there is some evidence to the contrary also.

7 Conclusion

This paper examines the role of vouchers as instruments of public school reform. It makes several important contributions in this context. First, it argues that all voucher programs are not created equal –

⁷³ <http://www.firn.edu/doe/sas/nrthome>

⁷⁴ Since the Stanford 9 test was first given in 2000, all these results are relative to the 2000 gains (which from the FCAT estimates are quite high.)

differences in voucher designs affect public school incentives differently and hence induce different responses from them. Therefore, understanding the effect of different voucher designs is essential to the formulation of effective voucher policies. This study contributes in this direction by comparing the effects of two U.S. voucher programs – Florida and Milwaukee – that differ fundamentally in their designs. In the context of an equilibrium theory of public school and household behavior, it argues that the Florida-type program should bring about an unambiguous improvement in public school performance and this improvement should exceed the improvement (if any) in the Milwaukee-type program. Using data from Florida and Milwaukee, and a difference-in-differences estimation strategy in trends, it then demonstrates that these findings are validated empirically. These findings are reasonably robust to alternative specifications and continue to hold after adjusting for mean-reversion.

Second, it also has important contributions from a theoretical point of view. It provides micro-foundations to the public school payoff function and derives the demand for public school from equilibrium household behavior. Moreover, it endogenously determines public school peer group quality at the respective program equilibria.

Third, empirical quantification, and even determining the direction, of public school response to vouchers is very difficult because it tends to be confounded with two other effects. Vouchers simultaneously affect the composition of the student body and peer quality – these too get reflected in public school quality or scores. The Florida program provides a unique case in this respect. In Florida, the schools that received an “F” grade in 1999 faced the threat of vouchers, and responded to it. However, if they managed to escape vouchers in the subsequent years of the program, the other two effects were absent. Therefore, the test scores of these schools provide a good source to assess the public school response.

An interesting direction of future research is to investigate whether these programs also induce public schools to manipulate the test taking population, or to focus on certain student groups at the expense of others. It may also be worth exploring the issue of “teaching to the test” in more detail. Understanding these issues will help design an effective voucher policy. I hope to address these issues in my future work.

Appendix A: Proofs of results

Proof of Proposition 1. This result can be proved in the following steps:

(i) Define $\Phi : [0, 1] \rightarrow [0, 1]$ such that for all $b' \in [0, 1]$, $b = \Phi(b') = \frac{\int_0^1 \int_0^{\hat{\alpha}(y, b', \cdot)} \alpha d\alpha dy}{\int_0^1 \int_0^{\hat{\alpha}(y, b', \cdot)} d\alpha dy}$.

Define a function F such that $F(\hat{\alpha}) = \frac{\int_0^1 \int_0^{\hat{\alpha}} \alpha d\alpha dy}{\int_0^1 \int_0^{\hat{\alpha}} d\alpha dy} = \Phi(b')$.

(ii) $\hat{\alpha}(y, b', \cdot)$ 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 $[0, 1] \rightarrow [0, 1]$.

(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. ■

Proof of $\frac{\delta g}{\delta b^e} > 0$:

Consider the sign of:

$$\int_0^1 (\hat{\alpha}(y, b^e, \cdot) - b) dy = \int_0^1 [\hat{\alpha}(y, b^e, \cdot) - \frac{1}{N(b^e, \cdot)} \cdot \int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha] dy \quad (\text{A.1})$$

The sign of $[\hat{\alpha}(y, b^e, \cdot) - \frac{1}{N(b^e, \cdot)} \cdot \int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha]$ will be the same as

$$[\hat{\alpha}(y, b^e, \cdot) \cdot N(b^e, \cdot) - \int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha] = \hat{\alpha}(y, b^e, \cdot) [\int_0^1 \hat{\alpha}(y, b^e, \cdot) dy - \frac{\hat{\alpha}(y, b^e, \cdot)}{2}].$$

This is positive since $\hat{\alpha}(y, b^e, \cdot) > 0$ and $\int_0^1 \hat{\alpha}(y, b^e, \cdot) dy - \frac{\hat{\alpha}(y, b^e, \cdot)}{2} > 0$. Therefore $\int_0^1 (\hat{\alpha}(y, b^e, \cdot) - b) dy > 0$. Intuitively, the positive sign can be seen from (A.1). For each y , $\hat{\alpha}(y, b^e, \cdot) > \frac{\int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha}{\int_0^{\hat{\alpha}(y, b^e, \cdot)} d\alpha} > \frac{\int_0^{\hat{\alpha}(y, b^e, \cdot)} \alpha d\alpha}{N(b^e, \cdot)}$. The first inequality follows because the highest public school ability at a certain income exceeds average of all abilities at that income. The second inequality follows because $N(b^e, \cdot) > \int_0^{\hat{\alpha}(y, b^e, \cdot)} d\alpha$. The former sums all public school households at all incomes while the latter sums public school households at only one income.

Now, $\frac{\delta^2 \hat{\alpha}}{\delta b^e \delta y} = -\frac{u_{b^e} \cdot (h_x^Q - h_x^2)}{(u(Q^*) - u(q))^2} < 0$ since $u_\theta > 0$, $h_{xx} < 0$ and $q_{b^e} > 0$. Therefore due to an increase in b^e

there is a higher increase in cutoff ability towards the lower incomes. Also $\hat{\alpha}(y)$ is inversely related to y .

Therefore higher positive values of $\frac{\delta \hat{\alpha}}{\delta b^e}$ are multiplied with higher positive values of $[\hat{\alpha}(y, b^e, \cdot) - b]$ so that

$\frac{\delta g}{\delta b^e}$ will be positive. Formally, if for some large y , $[\hat{\alpha}(y, b^e, \cdot) - b] < 0$, then there must exist some $y = y_1$

such that $\hat{\alpha}(y_1, b^e, \cdot) = b$. Then,

$$\begin{aligned} \int_0^{y_1} (\hat{\alpha}(y, b^e, \cdot) - b) dy &> | \int_{y_1}^1 (\hat{\alpha}(y, b^e, \cdot) - b^e) dy | \Rightarrow \int_0^{y_1} (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y_1, \cdot)}{\delta b^e} dy > | \int_{y_1}^1 (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y_1, \cdot)}{\delta b^e} dy | \\ &\Rightarrow \int_0^{y_1} (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta b^e} dy > \int_0^{y_1} (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y_1, \cdot)}{\delta b^e} dy > | \int_{y_1}^1 (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y_1, \cdot)}{\delta b^e} dy | > \\ &| \int_{y_1}^1 (\hat{\alpha}(y, b^e, \cdot) - b) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta b^e} dy | \end{aligned}$$

The last line follows because $\frac{\delta \hat{\alpha}}{\delta b^e}$ is positive and is strictly decreasing in y . Therefore $\frac{\delta g}{\delta b^e} > 0$.

Proof of Lemma 1.

$$b^* = \frac{1}{2} \cdot \frac{\int_0^1 \hat{\alpha}^2(y, b^*, \cdot) dy}{\int_0^1 \hat{\alpha}(y, b^*, \cdot) dy} = g(b^*, e, v, t, c) \text{ at equilibrium}$$

Effect of an increase in e on b^* :

$$\frac{\delta b^*}{\delta e} = \frac{\frac{\delta g(b^*, \cdot)}{\delta e}}{1 - \frac{\delta g(b^*, \cdot)}{\delta b}} \text{ where } \frac{\delta g(b^*, \cdot)}{\delta e} = \frac{1}{N(b^*, \cdot)} \cdot \int_0^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}}{\delta e} dy$$

The denominator is positive from uniqueness. Consider the numerator. $\frac{\delta \hat{\alpha}}{\delta e}$ is positive. Also,

$\frac{\delta^2 \hat{\alpha}}{\delta e \delta y} = -\frac{u_e \cdot (h_x^Q - h_x^q)}{(u(Q^*) - u(q))^2} < 0$ since $h_{xx} < 0$, $u_\theta > 0$ and $q_e > 0$. Therefore higher positive values of $\frac{\delta \hat{\alpha}}{\delta e}$ occur at lower values of y . Since $\hat{\alpha}(y)$ is inversely related to y , if for some high y , $(\hat{\alpha}(y, b^*, \cdot) - b^*) < 0$ then there must exist some y_2 such that $\hat{\alpha}(y_2, b^*, \cdot) = b^*$. From the proof of $\frac{\delta g}{\delta b^e} > 0$ above, $\int_0^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) dy > 0$.

Then it follows that:

$$\begin{aligned} \Rightarrow \int_0^{y_2} (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y_2, \cdot)}{\delta e} dy &> \left| \int_{y_2}^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y_2, \cdot)}{\delta e} dy \right| \text{ since } \frac{\delta \hat{\alpha}(y_2, \cdot)}{\delta e} \text{ is independent of } y. \\ &\Rightarrow \int_0^{y_2} (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta e} dy > \left| \int_{y_2}^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta e} dy \right| \text{ since } \frac{\delta^2 \hat{\alpha}}{\delta e \delta y} < 0 \end{aligned}$$

It follows that $\frac{\delta g}{\delta e} > 0$. Therefore $\frac{\delta b^*}{\delta e} > 0$.

Effect of an increase in v on b^* :

$$\frac{\delta b^*}{\delta v} = \frac{\frac{\delta g(b^*, \cdot)}{\delta v}}{1 - \frac{\delta g(b^*, \cdot)}{\delta b}} \text{ where } \frac{\delta g(b^*, \cdot)}{\delta v} = \frac{1}{N(b^*, \cdot)} \cdot \int_0^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}}{\delta v} dy$$

Consider the numerator. Now, $\frac{\delta^2 \hat{\alpha}}{\delta v \delta y} = \frac{-h_{xx}}{u(Q^*) - u(q)} > 0$. Therefore due to vouchers, there is a higher fall of cutoff ability towards the lower incomes. Since $\hat{\alpha}(y)$ is inversely related to y , higher positive values of

$\hat{\alpha}(y, b^*, \cdot) - b^*$ occur at lower incomes. Since $\frac{\delta \hat{\alpha}}{\delta v} < 0$, $\left| \int_0^{y_2} (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y_2, \cdot)}{\delta v} dy \right| > \int_{y_2}^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta v} dy \Rightarrow \left| \int_0^{y_2} (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta v} dy \right| > \int_{y_2}^1 (\hat{\alpha}(y, b^*, \cdot) - b^*) \cdot \frac{\delta \hat{\alpha}(y, \cdot)}{\delta v} dy$ since $\frac{\delta^2 \hat{\alpha}}{\delta v \delta y} > 0$

Therefore $\frac{\delta g}{\delta v} < 0$. This implies $\frac{\delta b^*}{\delta v} < 0$. ■

Proof of Lemma 2.

Equilibrium cutoff ability level at each income is given by $\hat{\alpha}(y; b^*, e, v, t, c)$.

$$\frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta e} = \frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta e} |b^* + \frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta b} \cdot \frac{\delta b^*}{\delta e} \quad \text{and} \quad \frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta v} = \frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta v} |b^* + \frac{\delta \hat{\alpha}(y; b^*, \cdot)}{\delta b} \cdot \frac{\delta b^*}{\delta v}.$$

Using lemma 1 and the signs of the partial derivatives, it follows that $\frac{\delta\hat{\alpha}(y;b^*,\cdot)}{\delta e} > 0$ and $\frac{\delta\hat{\alpha}(y;b^*,\cdot)}{\delta v} < 0$.

Symmetric analysis with the income cutoff at each ability level shows that $\frac{\delta\hat{y}(\alpha;b^*,\cdot)}{\delta e} > 0$ and $\frac{\delta\hat{y}(\alpha;b^*,\cdot)}{\delta v} < 0$.

■

Proof of Lemma 3.

Using the rent function, $\frac{\delta R(e,v)}{\delta v} = (p - c_N)N_v$. By assumption $(p - c_N) > 0$ and from proposition 2, $N_v < 0$. Therefore, $\frac{\delta R(e,v)}{\delta v} < 0$. ■

Proof of Proposition 3.

Under the VS program, e_{VS} solves the first order condition: $\frac{\delta R(e,v)}{\delta e} = (p - c_N)N_e(e, v) - C_e(e) = 0$.

Comparative statics with respect to v yields:

$$\frac{\delta e}{\delta v} = \frac{-[(p - c_N)N_{ev} - c_{NN}N_vN_e]}{(p - c_N)N_{ee} - c_{NN}N_e^2 - C_{ee}} \quad A.2$$

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

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

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

Proof of Proposition 4.

Proof to part (i):

Since $pN(e_{VS}, 0) - c(N(e_{VS}, 0)) - C(e_{VS}) > pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$

By the strict concavity of the rent function, $\exists \bar{e} > e_{VS}$ that satisfies the public school's incentive constraint with equality, $pN(\bar{e}, 0) - c(N(\bar{e}, 0)) - C(\bar{e}) = pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$.

Proof to part(ii): Facing the TOV program, if public school chooses not to meet the cutoff it anticipates vouchers, chooses the optimal effort e_{VS} and gets rent $pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$. If it chooses to meet the cutoff, it anticipates rent $pN(\bar{e}, 0) - c(N(\bar{e}, 0)) - C(\bar{e})$.

$pN(e_{PP}, 0) - c(N(e_{PP}, 0)) - C(e_{PP}) > pN(e_{VS}, 0) - c(N(e_{VS}, 0)) - C(e_{VS}) > pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS})$

The second inequality follows from lemma 3 and the first because e_{PP} is the rent maximizing effort under $v = 0$. Given strict concavity of the rent function, $\exists \bar{e} > e^{PP}$ such that it satisfies the public school's incentive constraint under TOV with equality,

$$pN(\bar{e}, 0) - c(N(\bar{e}, 0)) - C(\bar{e}) = pN(e_{VS}, v) - c(N(e_{VS}, v)) - C(e_{VS}). \quad \blacksquare$$

Proof of corollary 1.

Proof to part (ia): $e_{TOV} > e_{PP}$ from proposition 4 $\Rightarrow q(e_{TOV}, b^*(e_{TOV}, 0)) > q(e_{PP}, b^*(e_{PP}, 0))$ since $q_e > 0$, $q_b > 0$ and $\frac{\delta b^*}{\delta e} > 0$.

Proof to part (ib): $e_{TOV} > e_{VS}$ from proposition 4

$\Rightarrow q(e_{TOV}, b^*(e_{TOV}, 0)) > q(e_{VS}, b^*(e_{VS}, 0)) > q(e_{VS}, b^*(e_{VS}, v))$ since $q_e, q_b > 0$ and $\frac{\delta b^*}{\delta e} > 0$, $\frac{\delta b^*}{\delta v} < 0$.

Proof to part (ii):

$e_{VS} \geq e_{PP}$ from proposition 3 $\Rightarrow q(e_{VS}, b^*(e_{VS}, v)) < q(e_{VS}, b^*(e_{VS}, 0)) \geq q(e_{PP}, b^*(e_{PP}, 0))$. \blacksquare

Appendix B: Moral hazard problem – unobservable public school effort

This appendix relaxes the assumption of complete observability of public school effort⁷⁵ and examines whether under unobservable public school effort, the equilibrium effort under the TOV program still exceeds those under the PP and the VS programs. Given public school effort⁷⁶ $e \in [e_{min}, e_{max}]$, “effective effort” e' is realized according to the distribution $F(e'/e)$, where $e' \in [e'_{min}, e'_{max}]$. Although e is not publicly observable, all agents have complete knowledge of the set $[e_{min}, e_{max}]$ and the family of conditional distributions $F(e'/e)$ for $e \in [e_{min}, e_{max}]$. Public school quality ($q = q(e', b)$) is a composite of two factors: (i) “effective effort” e' and (ii) peer group quality (b) and can be thought of as being embodied in school scores. All agents observe quality q but not the actual public school effort e that generated it. The public school chooses effort e to maximize its expected rent. I make the following assumptions on the distribution function of e' .

A1. For each $e \in [e_{min}, e_{max}]$, $F(e'/e)$ admits a density $f(e'/e) > 0$ for all $e' \in [e'_{min}, e'_{max}]$. The density $f(e'/e)$ is twice continuously differentiable with respect to e .

A2. $f(e'/e)$ satisfies the strict monotone likelihood ratio property (MLRP) i.e for all $e, \hat{e} \in [e_{min}, e_{max}]$

⁷⁵ In practice, public school effort is not observable – what is observable is public school score which is a noisy representation of public school effort.

⁷⁶ Public School effort contains unobservable characteristics such as teacher and administrator effort as well as observable characteristics such as school facilities, class size, experience of teachers, school curriculum, after school programs, special education, etc. Presence of some unobservable characteristics makes the total e unobservable.

with $e > \hat{e}$, the likelihood ratio $\frac{f(e'/e)}{f(e'/\hat{e})}$ is strictly increasing in e' . This is equivalent to $\frac{\delta}{\delta e'} \left(\frac{f_e(e'/e)}{f(e'/e)} \right) > 0$ for all $e' \in [e'_{min}, e'_{max}]$. This assumption implies that $F(e'/e)$ is ordered according to first order stochastic dominance.

A3. $F(e'/e)$ satisfies the convexity of the distribution function condition (CDFC) i.e $F_{ee}(e'/e) > 0$ for all $e' \in [e'_{min}, e'_{max}]$ and $e \in [e_{min}, e_{max}]$.⁷⁷

The simple public-private system, the Milwaukee-type VS program and the Florida-type TOV program are modeled as follows. The public-private system has two stages. In the first stage, the public school chooses effort, which is not observable by the other agents of the economy. Quality q is realized and observed by all agents in the economy. In stage 2, households choose between schools. The VS program consists of three stages: In the first stage, the Government announces the voucher v . In stage 2, facing v , the public school chooses effort, which is not observable. Quality is realized and observed by all agents in the economy. In stage 3, households choose between schools. The Florida program is modeled in four stages: In the first stage, the Government announces the program and a cutoff quality level \bar{q} and voucher v . In stage 2, the public school chooses effort. Quality is realized and observed by all agents in the economy. In stage 3, government imposes vouchers if $q < \bar{q}$. No voucher is imposed if $q \geq \bar{q}$. In the final stage, households choose between schools.

The uncertainty signifies the absence of any direct one-to-one relationship between the effort of teachers and administrators and the realized “effective effort” and hence school scores. Each of the systems consists of a game between the public school and the households. Facing the program (if any) and correctly anticipating household behavior corresponding to all e' realizations (given e), the public school chooses effort to maximize its expected rent. Effective effort (e') is realized and given the existing peer quality,⁷⁸ school quality $q = q(e', b)$ is simultaneously realized and observed by all agents of the economy. Households observe this quality⁷⁹ and knowing existing peer quality deduce the “effective effort” of the school. Based on e' and whether vouchers were imposed, households anticipate a certain peer quality and make their utility maximizing choice. At equilibrium anticipated peer quality equals actual and a corresponding

⁷⁷ By the MLRP, $F(e'/e)$ decreases in e , CDFC requires that the function decrease at a decreasing rate. Together with the MLRP, the CDFC captures the idea that increasing effort also increases at a decreasing rate the probability $1 - F(e'/e)$ that the realization is greater than e' . Thus this is a form of stochastic diminishing returns to scale.

⁷⁸ The existing peer quality for both the VS and TOV systems can be thought of as the public-private equilibrium peer quality, since the latter is taken as the pre-program system for both programs. Nevertheless, it is not crucial what the existing peer quality is, but what is crucial is that all agents have complete knowledge of the existing peer quality.

⁷⁹ This quality q can be thought of as last year’s school scores.

equilibrium allocation of households between public and private schools is generated. The equilibria are defined exactly in the same way as earlier. For the sake of simplicity of notation, I still assume that $(e_{PP}, b_{PP}), (e_{VS}, b_{VS}), (e_{TOV}, b_{TOV})$ represent the equilibria under PP, VS and TOV systems.

7.0.1 Characterization of equilibria

Household behavior: Each household decides whether or not to go to public school based on “effective effort” e' and whether vouchers were imposed. A household (y, α) chooses private or public school according as $[u(y + v - t.Q^* - c) + \alpha h(Q^*)] - [u(y) + \alpha h(q(e', b))]$ is positive or negative. Thus behavior is basically the same as earlier, the only difference is that instead of using effort itself, they use a noisy representation of effort, effective effort e' to make their school choices. Therefore, all results in the last section follow through here – equilibrium peer quality, equilibrium cutoff ability level at each income and equilibrium number of public school students increase with e' and decrease with v .

Public School Behavior

The public school anticipates household behavior and chooses e to maximize expected rent:

$ER(v, \cdot) = \int_{e'_{min}}^{e'_{max}} [pN(e', v, b^*, \cdot) - c(N(e', v, b^*, \cdot))] f(e'/e) de' - C(e)^{80}$ where $v = 0$ under the public-private system. Equilibrium public school effort under the VS program can be either greater or less than the PP system⁸¹. The intuition behind this is as follows. With imposition of vouchers, rent decreases at each realization of e' .⁸² An increase in e increases the probability of higher e' realizations. However, the fall in rent with vouchers can either increase or decrease in e' .⁸³ This implies that vouchers may induce public schools to correspondingly decrease or increase effort in response to vouchers.

Under the Florida TOV program the public school faces a quality cutoff \bar{q} or equivalently an “effective effort” cutoff \bar{e}' and chooses e to maximize its expected rent. The school’s expected rent under the TOV

⁸⁰ The rent function is strictly concave. To see this, using integration by parts rewrite it as:

$$\{[pN(e', v, \cdot) - c(N(e', v, \cdot))]F(e'/e)\}_{e'_{min}}^{e'_{max}} - \int_{e'_{min}}^{e'_{max}} (p - c_N(\cdot))N_{e'}(\cdot)F(e'/e)de' - C(e)$$

$$= [pN(e'_{max}, v, \cdot) - c(N(e'_{max}, v, \cdot))] - \int_{e'_{min}}^{e'_{max}} (p - c_N(\cdot))N_{e'}(\cdot)F(e'/e)de' - C(e)$$
which is strictly concave since $F_{ee} > 0$ (CDFC), $C_{ee} > 0$.

⁸¹ To see this evaluate the first order condition under the VS system at e_{PP} :

$$\frac{\delta ER(v, \cdot)}{\delta e}|_{e_{PP}} = \frac{\delta ER(v, \cdot)}{\delta e}|_{e_{PP}} - \frac{\delta ER(0, \cdot)}{\delta e}|_{e_{PP}} = \int_{e'_{min}}^{e'_{max}} -\beta(e', v, \cdot)\gamma(e', e^{PP})f(e'/e^{PP})de'$$
where $\beta(e', v, \cdot) = \{[pN(e', 0, \cdot) - c(N(e', 0, \cdot))] - [pN(e', v, \cdot) - c(N(e', v, \cdot))]\}$. Now, $-\beta(e', v, \cdot) < 0$, but it can be either increasing or decreasing in e' . Therefore, although $\gamma(e', e^{PP})$ is always increasing and positive iff e' exceeds a certain cutoff, $\frac{\delta ER(v, \cdot)}{\delta e}|_{e_{PP}} \geq 0$.

⁸² Defining $r(e', v, \cdot) = pN(e', v, \cdot) - c(N(e', v, \cdot))$ for a certain realization e' , $r_v = \frac{\delta r}{\delta v} = (p - c_N(\cdot))N_v(\cdot) < 0$

⁸³ $r_{e'v} = \frac{\delta^2 r}{\delta e' \delta v} = (p - c_N)N_{e'v} - c_{NN}(\cdot)N_{e'}(\cdot)N_v(\cdot) \geq 0$ since $N_{e'v} \geq 0$, $N'_e > 0$, $c_{NN} > 0$, $p - c_N > 0$

program is given by:

$H = \int_{e'_{min}}^{\bar{e}'} [pN(e', v, \cdot) - c(N(e', v, \cdot))]f(e'/e)de' + \int_{\bar{e}'}^{e'_{max}} [pN(e', 0, \cdot) - c(N(e', 0, \cdot))]f(e'/e)de' - C(e)$. H is strictly concave in e.⁸⁴ The following lemma follows from the MLRP.

Lemma 4 *There exists $\hat{e}' \in (e'_{min}, e'_{max})$ such that $\gamma(e', e) \leq 0$ according as $e' \leq \hat{e}'$ where $\gamma(e', e) = \frac{f_e(e'/e)}{f(e'/e)}$.*

Proposition 5 *For all v , there exists a cutoff \bar{e}' ⁸⁵ such that the equilibrium effort under the TOV program exceeds those under (i) the public-private system and (ii) the voucher shock program.*

This proposition checks whether there exists any cutoff \bar{e}' such that the effort (e_{TOV}) induced by it exceeds those under the VS and PP equilibria. This is equivalent to finding the optimal contract (\bar{e}'^*, e_{TOV}^*) such that the cutoff \bar{e}'^* induces the maximum possible effort e_{TOV}^* given the public school incentive constraints and then checking whether e_{TOV}^* exceeds e_{VS} and e_{PP} . The PP and the VS scenarios can be incorporated under the broad cutoff design program or TOV program in the following way. If the cutoff is set at $\bar{e}' = e'_{max}$ the public school realizes that it will always get vouchers irrespective of what effort it chooses. So this regime is exactly the same as the “voucher shock” regime. The school’s best response here is therefore to choose e_{VS} , the optimum effort under the “voucher shock” regime. Similarly if $\bar{e}' = e'_{min}$ the public school realizes that it will never get vouchers no matter what effort it chooses. So this situation is equivalent to the pre-program public-private scenario and optimum effort of the school is e_{PP} , the optimum effort under the simple public-private system. This proposition shows that \bar{e}'^* lies strictly in the interior of (that is, $\bar{e}'^* \neq e'_{max}$ and $\bar{e}'^* \neq e'_{min}$) which implies that it must be the case that $e_{TOV}^* > e_{VS}$ and $e_{TOV}^* > e_{PP}$. However, it is possible to go a step further. The next proposition shows that there exists a range $[e'_1, e'_2] \subset [e'_{min}, e'_{max}]$ such that if $\bar{e}' \in [e'_1, e'_2]$ then the effort induced by \bar{e}' unambiguously exceeds e_{VS} and e_{PP} . Define E_1 and E_2 such that they respectively solve $\gamma(e', e_{VS}) = 0$ and $\gamma(e', e_{PP}) = 0$.

⁸⁴ Using integration by parts to rewrite H and noting that $C_{ee} > 0$ and $F_{ee} > 0$ it can be easily seen that $H_{ee} < 0$.

⁸⁵ Although the cutoff is considered in terms of “effective effort”, given the existing peer quality it can be easily translated in terms of quality.

Proposition 6 (i) *There exists $e'_1, e'_{min} \leq e'_1 < E_1$ such that if the cutoff $\bar{e}' \in [e'_1, e'_{max}]$ the effort under the “threat of voucher” program unambiguously exceeds that under the “voucher shock” program i.e., $e_{TOV} > e_{VS}$.*

(ii) *There exists $e'_2, E_2 < e'_2 \leq e'_{max}$ such that if the cutoff $\bar{e}' \in [e'_{min}, e'_2]$ the effort under the “threat of voucher” program unambiguously exceeds that under the pre-program public-private equilibrium i.e., $e_{TOV} > e_{PP}$.*

The intuitive argument behind this proposition is as follows. First consider the TOV and the VS programs. Facing the TOV program, if the school chooses e_{VS} (the equilibrium effort under the VS program), then at each realization of $e' < \bar{e}'$, its rent is the same as in the VS program. On the other hand, for each realization of $e' \geq \bar{e}'$, its rent is higher. Therefore the school chooses an effort strictly higher than e_{VS} to increase its probability of falling above \bar{e}' since it follows from the MLRP that an increase in effort increases the probability of higher e 's (lemma 4). The intuition behind $\bar{e} > e_{PP}$ is similar. Choosing e_{PP} under the TOV gives it the same rent as the PP program at each realization above \bar{e}' but lower rent at each realization below \bar{e}' . The school chooses an effort strictly above e_{PP} to increase (decrease) the probability of realizations above (below) \bar{e}' .⁸⁶ Thus, the results here parallel those in the complete information model (Proposition 4).

Appendix C: Proofs of Results in Appendix B

Proof of Lemma 4.

By the MLRP, $\gamma(e', e)$ is increasing in e' . Also for any $e, F(e'_{max}/e) = 1$. Therefore

$F_e(e'_{max}/e) = 0 \Rightarrow \int_{e'_{min}}^{e'_{max}} \gamma(e', e) f(e'/e) de' = 0$. Since $\gamma(e', e)$ is increasing in e' it follows that there exists some $\hat{e}' \in (e'_{min}, e'_{max})$ such that $\gamma(e', e) \leq 0$ according as $e' \leq \hat{e}'$. ■

Proof of Proposition 5.

Consider the following problem: $\max \int_{e'_{min}}^{e'_{max}} [e' + b(e', \cdot)] f(e'/e) de'$ ⁸⁷

$$\text{s.t. } \frac{\delta H}{\delta e} = \int_{e'_{min}}^{\bar{e}'} r(e', v) \frac{\delta f(e'/e)}{\delta e} dq + \int_{\bar{e}'}^{e'_{max}} r(e', 0) \frac{\delta f(e'/e)}{\delta e} dq - C_e(e) = 0$$

⁸⁶ However, although rent falls at each realization of e' with vouchers, this fall (or alternatively, the gain in rent from avoiding vouchers) may either increase or decrease with e' ($r_{e'v} \geq 0$). Depending on this, under certain circumstances, at very low levels of cutoff, the public school effort under TOV may be less than VS and at very high levels of cutoff, effort under TOV may be less than PP. The intuition is as follows. First consider TOV versus VS. If \bar{e}' is low, schools escape vouchers for low values of e' also. If it is the case that the gain in rent from avoiding vouchers is largest for lower values of e' then since an increase in effort decreases the probability of occurrence of lower values of e' , public school may find it profitable not to increase effort. Now consider TOV versus PP. If \bar{e}' is high, vouchers will be incurred at high values of e' also. If it is the case that the fall in rent due to vouchers is highest for high values of e' , then the school may not have an incentive to increase effort since an increase in effort increases the probability of occurrence of higher e' .

⁸⁷ Since effort improves e' in the sense of first order stochastic dominance and $[e' + b(e')]$ is increasing in e' the stochastic

If L denotes the Lagrangean and (\bar{e}', e_{TOV}^*) solves the above problem then there exists a non-zero multiplier λ such that $(\bar{e}', e_{TOV}^*, \lambda)$ solve the following three first order conditions:

$$\begin{aligned}\frac{\delta L}{\delta e'} &= \lambda \frac{\delta^2 H}{\delta e' \delta e} = \lambda [r(\bar{e}', v) - r(\bar{e}', 0)] f_e(\bar{e}'/e) = \lambda h_1(\bar{e}', e, \cdot) = 0 \\ \frac{\delta L}{\delta \lambda} &= \frac{\delta H}{\delta e} = \int_{e'_{min}}^{\bar{e}'} r(e', v) \frac{\delta f(e'/e)}{\delta e} de' + \int_{\bar{e}'}^{e'_{max}} r(e', 0) \frac{\delta f(e'/e)}{\delta e} de' - C_e(e) = 0 \\ \frac{\delta L}{\delta e} &= \int_{e'_{min}}^{e'_{max}} [e' + b(e')] f_e(e'/e) de' + \lambda [\int_{e'_{min}}^{\bar{e}'} r(e', v) f_{ee}(e'/e) \delta e' + \int_{\bar{e}'}^{e'_{max}} r(e', 0) f_{ee}(e'/e) \delta e' - C_{ee}(e)] = 0\end{aligned}$$

where $r(e', v) = pN(e', v) - c(N(e', v))$. The first two conditions determine \bar{e}'^* and e_{TOV}^* . λ is obtained as a solution to the third equation. Note that setting $\bar{e}' = e'_{max}$ and $\bar{e}' = e'_{min}$ is equivalent to the VS and PP regimes respectively. In the former case, the school gets vouchers always irrespective of the effort chosen, so this is the same as the VS regime and the school chooses the VS cutoff e_{VS} . In the latter case, it never gets vouchers, so this is equivalent to the PP regime and the school chooses e_{PP} . Therefore if it can be shown that $\bar{e}'^* \neq e'_{min}$ and $\bar{e}'^* \neq e'_{max}$ then it must generate effort $e_{TOV}^* > e_{PP}$ and $e_{TOV}^* > e_{VS}$.

Since $r_v(e', v) < 0$ and $\frac{\delta f(e'_{min}/e)}{\delta e} < 0$ (lemma 4) it follows that $\frac{\delta^2 H}{\delta e' \delta e}|_{e'_{min}} > 0$ that is, increasing \bar{e}' above e'_{min} increases effort. Therefore e'_{min} cannot be a solution to the problem. Similarly $\frac{\delta^2 H}{\delta e' \delta e}|_{e'_{max}} < 0$.

Therefore, e'_{max} cannot be a solution. Rewriting, $\frac{\delta^2 H}{\delta e' \delta e} = h_1(\bar{e}', e, v, t, c) = h_1(\bar{e}', e(\bar{e}'), v, t, c) = h_2(\bar{e}', v, t, c)$.

Since h_1 is continuous and the second FOC gives e as a continuous function of \bar{e}' , h_2 is continuous in \bar{e}' .

Therefore it follows that there exists an $\bar{e}'^* \in (e'_{min}, e'_{max})$ which solves the problem. ■

Proof of Proposition 6.

Proof to part (i): Evaluating the first order condition under the TOV program at e_{VS} :

$$\frac{\delta H}{\delta e}|_{e_{VS}} = \frac{\delta H}{\delta e}|_{e_{VS}} - \frac{\delta ER(v, \cdot)}{\delta e}|_{e_{VS}} = \int_{\bar{e}'}^{e'_{max}} [r(e', 0) - r(e', v)] f_e(e'/e_{VS}) de' = \int_{\bar{e}'}^{e'_{max}} \beta(e', v) f_e(e'/e_{VS}) de'$$

where $\beta(e', v) = [r(e', 0) - r(e', v)]$. From lemma 4, $f_e(e'/e_{VS}) \lesseqgtr 0$ according as $e' \lesseqgtr E_1$. Now if the cutoff $\bar{e}' \geq E_1$ then $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ since $\beta(e', v) > 0$ so that $\bar{e}' > e_{VS}$. There are two cases if $\bar{e}' < E_1$. Let

$\int_{e'_{min}}^{e'_{max}} \beta(e', v) f_e(e'/e_{VS}) de' = A_1$. Although $f_e(e'/e_{VS}) \lesseqgtr 0$ according as $e' \lesseqgtr E_1$, $\beta(e', v)$ may be increasing or decreasing in e' . Therefore $A_1 \gtrless 0$. (Note that $A_1 \leq 0$ implies $e_{PP} \leq e_{VS}$).

Case 1: If $A_1 > 0$ then for any $\bar{e}' \in (e'_{min}, E_1)$, $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ and $e_{TOV} > e_{VS}$.

distribution of e' induced by a higher effort is always preferred to one induced by a lower effort. This implies that this problem is equivalent to maximizing e subject to the school's incentive constraint.

Case 2: If $A_1 < 0$ then $\exists e'^1 \in (e'_{min}, E_1)$ such that $|\int_{e'^1}^{E_1} \beta(e', v) f_e(e'/e_{VS}) de'| = \int_{E_1}^{e'_{max}} \beta(e', v) f_e(e'/e_{VS}) de'$ then for any $\bar{e}' \in (e'^1, E_1)$, $\frac{\delta H}{\delta e}|_{e_{VS}} > 0$ and $e_{TOV} > e_{VS}$.

Using cases (1) and (2) define $e'_1 = [\min\{e' \in [e'_{min}, E_1] : \int_{e'}^{e'_{max}} \beta(e', v) f_e(e'/e_{VS}) de' > 0\}]$. Then for any $\bar{e}' \in [e'_1, e'_{max}]$, $e_{TOV} > e_{VS}$. Note that $e'_1 \geq e'_{min}$ according as $A_1 \leq 0$.

Proof to part (ii):

Evaluating the first order condition under the TOV program at e_{PP} :

$$\frac{\delta H}{\delta e}|_{e_{PP}} = \frac{\delta H}{\delta e}|_{e_{PP}} - \frac{\delta ER(0, \cdot)}{\delta e}|_{e_{PP}} = \int_{e'_{min}}^{\bar{e}'} [r(e', v) - r(e', 0)] f_e(e'/e_{PP}) de' = - \int_{e'_{min}}^{\bar{e}'} \beta(e', v) f_e(e'/e_{PP}) de'$$

From lemma 4, $f_e(e'/e_{PP}) \leq 0$ according as $e' \leq E_2$. Now if the cutoff $\bar{e}' \leq E_2$ then $\frac{\delta H}{\delta e}|_{e_{PP}} > 0$ so that $e_{TOV} > e_{PP}$. There are two cases if $\bar{e}' > E_2$. Let $A_2 = \int_{e'_{min}}^{e'_{max}} \beta(e', v) f_e(e'/e_{PP}) de'$. Again similarly as above $A_2 \geq 0$. (Note that $A_2 \geq 0$ implies $e_{PP} \geq e_{VS}$).

Case 1: If $A_2 < 0$ then for any $\bar{e}' \in (E_2, e'_{max})$ $\frac{\delta H}{\delta e}|_{e_{PP}} > 0$ and $e_{TOV} > e_{PP}$.

Case 2: If $A_2 > 0$ then $\exists e'^2 \in (E_2, e'_{max})$ such that $|\int_{e'^2}^{E_2} \beta(e', v) f_e(e'/e_{PP}) de'| = \int_{E_2}^{e'^2} \beta(e', v) f_e(e'/e_{PP}) de'$ then for any $\bar{e}' \in (E_2, e'^2)$, $\frac{\delta H}{\delta e}|_{e_{PP}} > 0$ and $e_{TOV} > e_{PP}$.

Using cases (1) and (2) define $e'_2 = [\max\{e' \in [E_2, e'_{max}] : \int_{e'_{min}}^{e'} -\beta(e', v) f_e(e'/e_{PP}) de' > 0\}]$. Then for any $\bar{e}' \in [e'_{min}, e'_2]$, $e_{TOV} > e_{PP}$. Note that $e'_2 \leq e'_{max}$ according as $A_2 \geq 0$. ■

References Angrist Joshua D., and Victor Lavy (1999), “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement,” *Quarterly Journal of Economics* 114(2), 533-575.

Chakrabarti, Rajashri (2003a), “Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee,” mimeo, Cornell University, May 2003.

Chakrabarti, Rajashri (2003b), “Can Increasing Private School Participation in a Voucher Program Affect Public School Performance? Evidence from the Milwaukee Voucher Experiment,” mimeo, Cornell University, October 2003.

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

Florida Department of State (1993-2003), “Florida Administrative Weekly,” Various Issues.

Greene, Jay P., Paul E. Peterson, Jiangtao Du, Leesa Boeger and Curtis L. Frazier (1996), “The effectiveness of School Choice in Milwaukee: A secondary Analysis of Data from the Program’s

Evaluation,” mimeo, University of Houston.

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

Hoxby, Caroline M. (2001), “School Choice and School Productivity (Or, Could School Choice be the tide that lifts all boats?)”, Harvard University typescript.

Hanushek, Eric A. and Margaret E. Raymond (2003), “Improving Educational Quality: How Best to Judge Our Schools?” in Yolanda Kodrzycki (ed.) *Education in the 21st Century: Meeting the Challenges of a Changing World*, Federal Reserve Bank of Boston, 193-224.

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.

Jacob, Brian A., and Lans Lefgren (2002), “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis,” NBER working paper #8918.

Laffont Jean-Jacques and David Martimort (2002), *The Theory of Incentives*, Princeton: Princeton University Press.

McMillan, Robert (2001), “Competition, Incentives, and Public School Productivity,” mimeo, University of Totonto.

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 Fiscal Bureau (1991-2003), “Milwaukee Parental Choice Program,” Informational Paper, Various Reports.

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

Table 1: Pre-program trend of F and D schools - testing the identifying assumptions

Florida FCAT Grade 4 Reading (1998-99), Grade 5 Math (1998-99) and Grade 4 Writing (1994-99)

	FCAT Reading		FCAT Math		FCAT Writing	
	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
trend	0.41 (0.56)	-0.05 (0.47)	13.20*** (0.55)	13.02** (0.61)	0.20** (0.008)	0.21** (0.003)
treated * trend	-1.78 (2.47)	-2.01 (1.46)	-0.98 (1.44)	-0.72 (1.48)	-0.05*** (0.011)	-0.04*** (0.007)
Controls	Y	Y	Y	Y	Y	Y
Observations	1013	1013	1006	1006	2948	2948
R-squared	0.58	0.93	0.59	0.90	0.64	0.80

*, **, ***: significant at the 10, 5, and 1 percent level, respectively.

Huber-White standard errors are in parenthesis. All regressions are weighted by the number of students tested. The OLS columns allow for correlations within districts. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 2: Effect of “Threatened Status” on FCAT Grade 4 Reading Scores

(Sample of treated F and control D schools in Florida, 1998-2002)

	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
trend	-0.59 (0.57)	-0.20 (0.67)				
Program dummy	-5.25*** (1.46)	-5.30*** (0.84)				
Treated * Program dummy	2.71* (1.60)	2.97* (1.78)	2.74* (1.65)	2.99* (1.80)		
Program dummy * trend	5.35*** (0.57)	5.57*** (0.76)				
Treated * Program dummy * trend	1.57** (0.74)	1.10 (1.00)	1.59** (0.76)	1.09 (1.00)		
Treated * 1 year after program					4.85*** (1.49)	4.85*** (1.68)
Treated * 2 years after program					4.71*** (1.78)	3.30* (1.71)
Treated * 3 years after program					8.01*** (1.49)	7.08*** (1.78)
Controls	N	Y	N	Y	N	Y
Year dummies	N	N	Y	Y	Y	Y
Observations	2567	2550	2567	2550	2567	2550
R-squared	0.11	0.77	0.11	0.77	0.11	0.77

*, **, ***: significant at the 10, 5, and 1 percent level, respectively.

Huber-White standard errors are in parenthesis. All regressions are weighted by the number of students tested. The OLS columns allow for correlations within districts. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 3: Effect of “Threatened Status” on FCAT Grade 5 Math (1998-2002) and Grade 4 Writing Scores (1994-2002)

(Sample of treated F and control D schools in Florida)

	Math						Writing					
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
trend	12.35*** (0.66)	12.80*** (0.76)					0.20*** (0.01)	0.21*** (0.004)				
Program dummy	0.01 (0.86)	0.31 (0.91)					0.11*** (0.03)	0.10*** (0.02)				
Treated * Program dummy	7.95*** (2.18)	7.90** (2.29)	7.93*** (2.19)	7.88*** (2.30)			0.30*** (0.09)	0.30*** (0.05)	0.30*** (0.09)	0.31*** (0.05)		
Program dummy * trend	-8.90 (0.72)	-9.11*** (0.85)					-0.13*** (0.01)	-0.13*** (0.01)				
Treated * Program dummy * trend	-0.51 (1.06)	-0.71 (1.04)	-0.50 (1.06)	-0.60 (1.06)			0.05** (0.02)	0.04** (0.02)	0.05*** (0.02)	0.04** (0.02)		
Treated * 1 year after program					7.18*** (1.94)	6.78*** (1.63)					0.35*** (0.08)	0.35*** (0.04)
Treated * 2 years after program					7.45*** (2.23)	7.25*** (1.82)					0.39*** (0.08)	0.37*** (0.04)
Treated * 3 years after program					6.17*** (3.06)	5.35*** (2.00)					0.44*** (0.06)	0.43*** (0.05)
Year dummies	N	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Observations	2541	2524	2541	2524	2541	2524	4513	4476	4513	4476	4513	4476
R-squared	0.19	0.76	0.19	0.76	0.19	0.76	0.73	0.84	0.75	0.85	0.75	0.85
p-value ¹	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. ¹ p-value of F-test of the program effect on treated schools. Huber-White standard errors are in parenthesis. All regressions are weighted by the number of students tested. The OLS columns allow for correlations within districts. Columns (1)-(6) include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure as controls. Columns (7)-(12) include an interaction term of treated dummy with trend and control for race, sex and per pupil expenditure.

Table 4: Effect of “Threatened Status” on FCAT Grade 4 Reading Scores – Looking for Post-Program intercept, trend and non-linear year shifts
(Sample of treated F, D and control C schools in Florida, 1998-2002)

	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Trend	2.49 (0.58)	2.77 (0.43)				
Program dummy	-7.00*** (1.45)	-6.46*** (0.54)				
Program dummy * trend	0.02 (0.93)	0.22 (0.48)				
Less treated * program dummy	1.70** (0.84)	1.20 (1.00)	1.73** (0.84)	1.25* (0.99)		
More treated* program dummy	4.88** (2.04)	4.99** (2.43)	4.90** (2.03)	5.07** (2.44)		
Less Treated * program dummy * trend	5.22*** (0.81)	4.85*** (0.93)	5.21*** (0.81)	4.83*** (0.93)		
More treated*program dummy*trend	7.73*** (2.47)	8.02*** (2.08)	7.71*** (2.48)	8.01*** (2.09)		
Less treated * 1 year after program					4.13*** (0.75)	3.53*** (0.76)
Less treated * 2 years after program					5.84*** (0.96)	5.52*** (0.80)
Less treated * 3 years after program					8.60*** (0.93)	7.94*** (0.87)
More treated * 1 year after program					9.45***††† (1.92)	9.32***††† (1.87)
More treated * 2 years after program					11.01***†† (2.39)	10.75***†† (1.87)
More treated * 3 years after program					17.08***††† (2.64)	16.03***††† (1.91)
Controls	N	Y	N	Y	N	Y
Year dummies	N	N	Y	Y	Y	Y
Observations	6034	5933	6034	5933	6034	5933
R-squared	0.44	0.86	0.44	0.86	0.44	0.86
p-value ¹	0.00	0.00	0.00	0.00	0.00	0.00

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. †††: more treated significantly different from less treated at 10%, 5% and 1% respectively. ¹ p-value of the F-test for joint significance of the post-program shifts of the more treated schools. Huber-White standard errors are in parenthesis. The OLS columns allow for correlations within districts. Columns (1)-(4) include interactions of trend with more treated and less treated dummies respectively. Columns (5)-(6) include interactions of D_1 dummy ($D_1 = 1$ if year > 1998) with more treated and less treated dummies respectively. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 5: Effect of “Threatened Status” on FCAT Grade 5 Math (1998-2002) and Grade 4 Writing Scores (1994-2002)

(Sample of treated F, D and control C schools in Florida)

	Math					Writing				
	OLS	FE	FE	OLS	FE	OLS	FE	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Trend	9.82*** (0.45)	10.20*** (0.48)				0.19*** (0.004)	0.19*** (0.002)			
Program dummy	0.16 (0.86)	0.60 (0.56)				0.03** (0.03)	0.02** (0.01)			
Program dummy * trend	-7.78*** (0.60)	-7.80 (0.53)				-0.09*** (0.01)	-0.09*** (0.01)			
Less treated * program dummy	-0.23 (0.92)	-0.15 (1.09)	-0.17 (1.09)			0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)		
More treated * program dummy	9.05***††† (2.13)	9.02***††† (2.23)	8.96***††† (2.24)			0.38***††† (0.09)	0.37***††† (0.05)	0.38***††† (0.05)		
Less treated*program dummy *trend	0.29 (0.60)	0.30 (1.04)	0.49 (1.03)			-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)		
More treated*program dummy *trend	0.91 (1.10)	0.54 (1.01)	0.50 (1.02)			0.01††† (0.02)	0.00† (0.02)	0.00† (0.02)		
Less treated * 1 year after prog				1.06*** (0.73)	0.97 (0.85)				0.05** (0.02)	0.05** (0.02)
Less treated * 2 years after prog				2.83** (1.44)	2.54*** (0.94)				0.00 (0.03)	0.00 (0.02)
Less treated * 3 years after prog				3.89*** (1.20)	3.47*** (0.92)				-0.03 (0.04)	-0.03 (0.02)
More treated * 1 year after prog				9.56***††† (2.02)	8.96***††† (1.59)				0.40***††† (0.07)	0.39***††† (0.04)
More treated * 2 years after prog				11.61***††† (2.99)	11.00***††† (1.77)				0.39***†† (0.07)	0.37***†† (0.04)
More treated * 3 years after prog				11.39***††† (3.46)	11.94***††† (1.95)				0.42***†† (0.05)	0.39***†† (0.05)
Year dummies	N	N	Y	Y	Y	N	N	Y	Y	Y
Observations	6003	5909	5909	6003	5909	10646	10587	10587	10646	10587
R-squared	0.44	0.83	0.83	0.44	0.83	0.72	0.85	0.86	0.74	0.86
p-value ¹	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

¹ p-value of the F-test of program effect on treated schools. Huber-White standard errors are in parenthesis. The OLS columns allow for correlations within districts. All regressions are weighted by the number of students tested. Columns (1)-(3) include interaction between trend and less treated dummy, (4)-(5) interaction between D_1 dummy ($D_1 = 1$ if year > 1998) and less treated dummy, (6)-(10) interactions between trend and less treated and more treated dummies respectively. The FE columns include school fixed effects and control for race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 6: Effect of “Voucher Shock” on Treatment Status

WRCT Grade 3 Reading % above and % below Scores (1989-1997), ITBS Grade 5 Reading (1987-93) and ITBS Grade 5 Math (1987-97)

	WRCT (% above)			WRCT (% below)			ITBS Reading			ITBS Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Less treated * program dummy	3.45 (2.98)	2.01 (3.01)		-2.02 (2.25)	-0.15 (2.30)							
Somewhat treated*program dummy	3.50 (2.59)	3.17 (2.50)		-3.72* (1.94)	-2.87 (1.91)		3.57 (5.24)	3.61 (5.22)		0.39 (2.81)	0.63 (2.82)	
More treated * program dummy	2.85 (3.32)	2.74 (3.15)		-1.60 (2.56)	-1.67 (2.50)		1.61 (4.98)	1.86 (5.01)		-2.97 (3.13)	-2.16 (3.22)	
Less treated*program dummy*trend	0.26 (0.53)	-0.57 (0.51)		0.00 (0.38)	-0.11 (0.37)							
Somewhat treated*program dummy *trend	0.64 (0.47)	0.70 (0.44)		-0.26 (0.34)	-0.33 (0.33)		1.34 (1.94)	1.34 (1.96)		0.61 (0.54)	0.54 (0.56)	
More treated*program dummy*trend	0.67 (0.62)	0.92 (0.58)		0.14 (0.46)	0.05 (0.46)		1.47 (2.31)	1.45 (3.30)		0.50 (0.63)	0.55 (0.71)	
Less treated * 1 year after program			3.28 (3.19)			0.94 (2.06)						
Less treated * 2 years after program			3.51 (2.67)			-0.79 (1.85)						
Less treated * 3 years after program			4.25 (3.74)			0.86 (2.52)						
Somewhat treated * 1 year after program			2.03 (2.81)			-0.54 (2.05)		4.15 (4.49)				-1.35 (2.94)
Somewhat treated * 2 years after program			5.38** (2.43)			-4.45* (1.88)		7.83 (5.17)				6.14* (3.38)
Somewhat treated * 3 years after program			5.01 (3.03)			-2.60 (2.30)		6.78 (5.31)				2.47 (3.31)
More treated * 1 year after program			-0.92 (3.33)			1.55 (2.50)		1.12 (3.86)				-4.02 (3.26)
More treated * 2 years after program			6.06* (3.14)			-4.16* (2.50)		6.59 (5.15)				4.36 (3.83)
More treated * 3 years after program			5.69 (3.98)			0.38 (3.16)		2.85 (5.18)				-2.22 (3.54)
Year dummies	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
Observations	1195	1195	1195	1195	1195	1195	717	717	717	1127	1127	1127
R-squared	0.50	0.57	0.58	0.47	0.52	0.52	0.55	0.55	0.55	0.58	0.59	0.60
p-value ¹	0.06	0.06	0.11	0.82	0.75	0.11	0.67	0.67	0.62	0.49	0.76	0.27

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. ¹ p-value of the F-test of joint significance of more treated shift coefficients. Huber-White standard errors are in parenthesis. This table uses the 66-47 sample. All regressions include school fixed effects and control for ethnicity, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 7: Effect of Milwaukee “Voucher Shock” Program using a continuous treatment variable

(Sample of Milwaukee Public Schools)

	Dependent variable: Grade 3 WRCT reading (% above) Scores					
	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
trend	-8.36*** (1.39)	-8.32*** (1.55)				
Program dummy	-1.68 (5.49)	-0.08 (4.61)				
Treatment * Program dummy	0.11 (0.09)	0.09 (0.08)	0.11 (0.09)	0.01 (0.09)		
Program dummy * trend	8.40*** (1.80)	9.77*** (1.86)				
Treatment * Program dummy * trend	0.02 (0.02)	0.00 (0.02)	0.01 (0.02)	0.00 (0.01)		
Treatment * 1 year after program					0.00 (0.09)	-0.09 (0.09)
Treatment * 2 years after program					0.19** (0.09)	0.07 (0.09)
Treatment * 3 years after program					0.14 (0.12)	0.03 (0.11)
Controls	N	Y	N	Y	N	Y
Year dummies	N	N	N	N	Y	Y
Observations	920	920	920	920	920	920
R-squared	0.12	0.47	0.20	0.55	0.20	0.55

*, **, ***: significant at the 10, 5, and 1 percent level, respectively.

Huber-White standard errors are in parenthesis. Controls include race, sex and percentage of students eligible for free or reduced-price lunches.

Table 8: Comparing the impact of the Florida “threat of voucher” program and the Milwaukee “voucher shock” program

Using performance in reading test [WRCT (% above) 1989-97 and FCAT Reading 1998-2002] and math test [ITBS Math 1986-97 and FCAT Math 1998-2002]

	Reading						Math			
	Milwaukee		Florida	Milwaukee		Florida	Milwaukee	Florida	Milwaukee	Florida
	WRCT	ITBS	FCAT	WRCT	ITBS	FCAT	ITBS	FCAT	ITBS	FCAT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
More Treated * Program dummy	0.18	0.09	0.25***				-0.18	0.45***		
More Treated * Program dummy * trend	0.04	0.08	0.40***				0.03	0.03		
More Treated * 1 year after prog				-0.06	0.06	0.47***			-0.24	0.45***
More treated * 2 years after prog				0.38*	0.35	0.50***			0.26	0.55***
More treated * 3 years after prog				0.35	0.15	0.80***			-0.13	0.60***
trend	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No
Post-program common trend & intercept shifts	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No
Year dummies	No	No	No	Yes	Yes	Yes	No	No	Yes	Yes

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. All figures are in terms of respective sample standard deviations. This table uses the 66-47 sample for Milwaukee. All figures are obtained from regressions that contain ethnicity, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure as control variables. Standard deviation of FCAT reading scores = 20, Standard deviation of FCAT math scores = 20, Standard deviation of WRCT (% above) reading scores = 16, Standard deviation of ITBS reading scores = 18.45, Standard deviation of ITBS math scores = 16.71

Table 9: Mean Reversion of the 98F schools compared to 98D and 98C schools, 1998-1999.

Panel A: 98F and 98D Schools		Dependent Variable: FCAT Score, 1998-99.					
	Reading		Math		Writing		
	(1)	(2)	(3)	(4)	(5)	(6)	
	OLS	FE	OLS	FE	OLS	FE	
trend	2.27*** (0.67)	2.01*** (0.43)	14.25*** (0.65)	14.02*** (0.49)	0.04*** (0.01)	0.04*** (0.01)	
98F*trend	-0.45 (1.77)	-0.65 (1.14)	1.03 (1.81)	1.17 (1.19)	0.14*** (0.03)	0.14*** (0.02)	
Observations	1353	1353	1354	1354	1355	1355	
R ²	0.64	0.93	0.63	0.91	0.33	0.85	

Panel B: 98F, 98D and 98C Schools		Dependent Variable: FCAT Score, 1998-99.					
	Reading		Math		Writing		
	(1)	(2)	(3)	(4)	(5)	(6)	
	OLS	FE	OLS	FE	OLS	FE	
trend	1.76** (0.56)	1.76*** (0.35)	9.57*** (0.50)	9.71*** (0.36)	0.03*** (0.01)	0.03*** (0.01)	
98F*trend	0.18 (1.78)	-0.55 (1.12)	4.67*** (1.80)	4.63*** (1.16)	0.14*** (0.03)	0.14*** (0.02)	
98D*trend	0.41 (0.88)	0.16 (0.54)	4.61*** (0.82)	4.22*** (0.58)	0.01 (0.02)	0.01 (0.01)	
Observations	2605	2605	2608	2608	2608	2608	
R ²	0.76	0.96	0.76	0.94	0.38	0.87	

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. Huber-White standard errors are in parenthesis. All regressions include race, sex, percentage of students eligible for free or reduced-price lunches and per pupil expenditure as controls. Sample of 98F and 98D schools: s.d of FCAT reading, math and writing respectively are 18.9, 18.05, 0.30. Sample of 98F, 98D, 98C schools: s.d of FCAT reading, math and writing respectively are 21.16, 21.56 and 0.31.

Table 10: Comparing the impact of the Florida threat of voucher program and Milwaukee voucher shock program, after correcting for mean reversion

Using performance in reading test [WRCT (% above) 1989-97 and FCAT Reading 1998-2002] and math test [ITBS Math 1986-1997 and FCAT Math 1998-2002]

	Corrected for Mean Reversion							
	Reading		Math		Reading		Math	
	Wisconsin WRCT (1)	Florida FCAT (2)	Wisconsin ITBS (3)	Florida FCAT (4)	Wisconsin WRCT (5)	Florida FCAT (6)	Wisconsin ITBS (7)	Florida FCAT (8)
More Treated * 1 year after prog	-0.06	0.47***	-0.24	0.45***	-0.06	0.47***	-0.24	0.24***
More treated * 2 years after prog	0.38*	0.50 ***	0.26	0.55***	0.38*	0.50***	0.26	0.34***
More treated * 3 years after prog	0.36	0.80***	-0.13	0.60***	0.36	0.80***	-0.13	0.39***
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All figures are in terms of respective sample standard deviations. The insignificant coefficients are reported as zero standard deviation. All figures are obtained from regressions that contain ethnicity, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure as control variables.

Table 11: Pre-program summary characteristics of F and D Schools in Florida, 1999

	Sample of all F and D Schools			Discontinuity Sample 1		
	F (std. dev.)	D (std.dev.)	F-D [p-value]	F (std. dev.)	D (std.dev.)	F-D [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)
% black	62.79 (28.23)	46.56 (30.98)	16.23 [0.00]	57.97 (29.45)	56.36 (30.02)	1.60 [0.77]
% hispanic	18.95 (23.40)	19.31 (24.10)	-0.37 [0.91]	17.61 (24.22)	18.95 (22.56)	-1.34 [0.74]
% white	17.18 (19.55)	32.20 (28.21)	-15.02 [0.00]	23.18 (21.34)	23.11 (25.28)	0.06 [0.99]
%asian	0.61 (1.19)	1.19 (2.39)	-0.58 [0.05]	0.73 (1.25)	0.99 (2.48)	-0.25 [0.54]
% american indian	0.22 (0.59)	0.25 (0.75)	-0.03 [0.73]	0.27 (0.68)	0.24 (0.80)	0.03 [0.83]
%male	51.38 (4.84)	51.07 (5.86)	-0.32 [0.68]	0.73 (1.25)	0.99 (2.48)	-0.25 [0.54]
% free-reduced lunch	85.80 (9.95)	74.78 (16.39)	-11.02 [0.00]	83.97 (8.39)	81.46 (12.47)	2.51 [0.23]
FCAT Reading Score	253.55 (17.68)	272.34 (17.12)	-18.79 [0.00]	263.17 (12.04)	262.94 (15.37)	0.23 [0.93]
# tested in Reading	87.51 (29.37)	90.18 (40.86)	-2.67 [0.61]	90.13 (32.43)	89.65 (38.45)	0.48 [0.94]
FCAT Math Score	274.31 (13.99)	290.30 (14.86)	-15.99 [0.00]	281.64 (9.09)	282.65 (12.69)	-1.01 [0.64]
# tested in Math	82.66 (25.94)	86.41 (39.57)	-3.75 [0.46]	85.97 (30.14)	85.18 (37.60)	0.79 [0.90]
FCAT Writing Score	2.45 (0.17)	2.85 (0.27)	-0.40 [0.00]	2.51 (0.14)	2.72 (0.20)	-0.21 [0.00]
# tested in Writing	89.09 29.61	90.42 (40.89)	-1.33 [0.80]	92.26 (32.48)	90.28 (38.38)	1.98 [0.77]
# of Schools	65	457		39	173	

Table 12A: Regression Discontinuity Analysis in Florida
Effect of “Threatened Status” on FCAT Grade 4 Reading (1998-2002), Writing (1994-2002) and
grade 5 Math (1998-2002) Scores

Panel A	Reading		Math		Writing	
	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity Sample 1						
trend	-0.56 (1.10)		12.91*** (1.22)		0.20*** (0.01)	
Program dummy	-6.47*** (1.42)		-2.03 (1.65)		0.16*** (0.03)	
Program dummy * trend	6.74*** (1.26)		-8.91*** (1.39)		-0.12 (0.01)	
Treated*Program dummy	4.05* (2.16)		10.29*** (3.06)		0.17 (0.07)	
Treated*Program dummy*trend	0.75 (1.40)		-1.65 (1.43)		0.05 (0.03)	
Treated*1 year after program		4.76** (2.23)		8.64*** (2.15)		0.23*** (0.05)
Treated*2 years after program		2.55 (2.21)		(6.83)*** (2.51)		0.25*** (0.05)
Treated*3 years after program		6.31** (2.52)		5.33*** (2.80)		0.34*** (0.06)
Controls	Y	Y	Y	Y	Y	Y
Observations	1043	1043	1033	1033	1847	1847
R-squared	0.70	0.62	0.67	0.68	0.84	0.85
p-value	0.03	0.04	0.00	0.00	0.00	0.00
Panel B	Discontinuity Sample 2			Discontinuity Sample 3		
	Reading	Math	Writing	Reading	Math	Writing
	FE	FE	FE	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Treated * Program dummy	5.84* (3.52)	10.92*** (3.81)	0.16** (0.08)	6.65* (4.00)	10.09** (4.56)	0.16* (0.09)
Treated * Program dummy * trend	-0.83 (1.70)	0.12 (1.76)	0.06* (0.04)	-1.65 (4.11)	2.66 (1.90)	(0.04) (0.04)
Controls	Y	Y	Y	Y	Y	Y
Observations	563	573	1016	341	331	613
R-squared	0.79	0.75	0.85	0.78	0.72	0.84
p-value	0.01	0.00	0.00	0.08	0.00	0.00
# of F Schools (% in group)	27 (40)	27 (40)	27 (40)	21 (20)	21 (20)	21 (20)
# of D Schools (% in group)	90 (20)	90 (20)	90 (20)	47 (10)	47 (10)	47 (10)

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. Huber-White standard errors are in parenthesis. All regressions are weighted by the number of students tested. The writing columns control for pre-program differences in trend or year effects of treated schools. All regressions include school fixed effects. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure. All columns in Panel B include trend, program dummy and interaction between program dummy and trend.

Table 12B. Regression Discontinuity Analysis in Florida (Continued)

	Discontinuity Sample 4		
	Reading FE (1)	Math FE (2)	Writing FE (3)
Treated * 1 year after Program	4.21* (2.46)	8.03*** (2.58)	0.19*** (0.06)
Treated * 2 years after Program	3.45 (2.71)	7.12** (3.04)	0.12* (0.06)
Treated * 3 years after Program	7.47** (3.06)	6.49** (3.26)	0.20*** (0.08)
Controls	Y	Y	Y
Observations	513	505	909
R-squared	0.76	0.76	0.87
p-value	0.01	0.00	0.00

*, **, *** denote significance at the 10, 5, and 1 percent levels respectively. The table uses sample of F and D schools where both groups fail to achieve the minimum criteria in both reading and mathematics - in addition, the F schools, but not the D schools, fail to achieve the minimum criterion in writing.

**Table 13. Is there a stigma effect of getting the lowest performing grade?
Effect of being categorized in group 1 on FCAT Writing Scores**

	Using FCAT Writing Scores, 1997-1998					
	Sample of Group 1 and 2 Schools			Sample of Group 1, 2 and 3 Schools		
	OLS (1)	FE (2)	FE (3)	OLS (4)	FE (5)	FE (6)
Trend	0.52*** (0.04)	0.52*** (0.03)	0.48*** (0.04)	0.48*** (0.02)	0.48*** (0.01)	0.46*** (0.02)
Group 1*trend	-0.01 (0.08)	-0.02 (0.06)	-0.02 (0.06)	0.03 (0.07)	0.01 (0.05)	0.02 (0.05)
Group2*trend				0.03 (0.04)	0.04 (0.03)	0.04 (0.03)
Controls	N	N	Y	N	N	Y
Observations	314	314	314	1361	1361	1358
R-squared	0.49	0.84	0.85	0.52	0.87	0.87

*, **, ***: significant at the 10, 5, and 1 percent level, respectively. Huber-White standard errors are in parenthesis. Controls include race, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table 14: Has there been “teaching to the test” in Florida?

Panel A	Correlation between FCAT and Stanford 9 NPR Scores			
	All Schools	F Schools	D Schools	C Schools
Grade 4 Reading, 2000	0.94 (1603)	0.85 (65)	0.87 (453)	0.85 (695)
Grade 4 Reading, 2001	0.96 (1651)	0.81 (64)	0.91 (451)	0.92 (694)
Grade 4 Reading, 2002	0.95 (1694)	0.85 (63)	0.88 (451)	0.91 (694)
Grade 5 Math, 2000	0.93 (1599)	0.81 (65)	0.85 (452)	0.84 (694)
Grade 5 Math, 2001	0.95 (1650)	0.89 (64)	0.91 (448)	0.91 (692)
Grade 5 Math, 2002	0.95 (1699)	0.85 (63)	0.90 (447)	0.92 (694)
Change in Grade 4 Reading, 2001-00	0.92	0.91	0.91	0.92
Change in Grade 4 Reading, 2002-01	0.94	0.92	0.93	0.95
Change in Grade 5 Math, 2001-00	0.93	0.93	0.92	0.92
Change in Grade 5 Math, 2002-01	0.95	0.94	0.94	0.94

Panel B	Dependent Variable = Stanford 9 NPR Scores					
	Reading			Math		
	Grade 3 (1)	Grade 4 (2)	Grade 5 (3)	Grade 3 (4)	Grade 4 (5)	Grade 5 (6)
More treated * 2001 year dummy	1.53*** (0.83)	3.41*** (0.82)	2.19*** (0.71)	3.08*** (1.00)	3.35*** (1.01)	3.06*** (1.08)
More treated * 2002 year dummy	2.58*** (0.91)	4.40***††† (0.89)	2.12*** (0.72)	4.62***† (1.12)	6.33***††† (1.09)	1.87* (1.10)
Less treated * 2001 year dummy	1.49*** (0.42)	2.59*** (0.41)	2.31*** (0.39)	1.75*** (0.48)	2.84*** (0.44)	2.45*** (0.49)
Less treated * 2002 year dummy	2.45*** (0.44)	2.49*** (0.43)	2.96 (0.42)	2.22*** (0.52)	3.24*** (0.47)	2.90*** (0.50)
Observations	3546	3545	3530	3546	3544	3530
R ²	0.91	0.91	0.93	0.90	0.92	0.89
p-value	0.02	0.00	0.01	0.00	0.00	0.02

Panel 1: All correlations in are significantly different from zero at the 1% level. The number of schools in the corresponding category are in parentheses.

Panel 2: *, **, ***: significant at the 10, 5, and 1 percent level, respectively. ¹p-value of the F-test for joint significance of post-program more treated year effects. Huber-White standard errors are in parenthesis. All regressions are weighted by the number of students tested and include school fixed effects ethnicity, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Table B.1: Pre-program demographic characteristics of more treated schools and control schools in Florida and Wisconsin

Panel A	Florida	Wisconsin		Florida–Wisconsin	
		66-47	60-47	66-47	60-47
More treated Schools	(std. dev.)	(std.dev.)	(std.dev.)	[p-value]	[p-value]
% black	62.79 (28.23)	66.55 (32.22)	62.90 (29.58)	-3.76 [0.56]	-0.10 [0.99]
% hispanic	18.95 (23.40)	18.07 (24.54)	14.81 (21.86)	0.88 [0.87]	4.14 [0.36]
% white	17.18 (19.54)	10.21 (10.68)	17.38 (16.55)	6.97 [0.07]	-0.20 [0.96]
%male	51.38 (4.84)	52.25 (2.60)	52.33 (2.58)	-0.87 [0.34]	-0.95 [0.22]
% free-reduced lunch	85.80 (9.95)	84.5 (6.48)	82.9 (9.04)	1.3 [0.50]	2.9 [0.12]

Panel B	Florida	Wisconsin	Florida–Wisconsin
Control Schools	(std.dev.)	(std. dev.)	[p-value]
% black	18.12 (14.17)	22.37 (12.93)	-4.25 [0.10]
% hispanic	15.49 (21.23)	14.84 (6.02)	0.17 [0.86]
% white	63.59 (22.33)	60.85 (12.80)	2.73 [0.49]
% male	51.38 (4.84)	50.63 (2.29)	0.76 [0.43]
% free-reduced lunch	50.14 (17.51)	44.95 (11.66)	5.19 [0.10]

Table B.2: Percent of Schools in Bottom 10th and 25th Percentile in FCAT in 1998 & 1999

Panel A: # of years	% of Schools in Bottom 10th percentile				
	Certainty Case	Lottery Case	Reading	Math	Writing
during 1998 & 1999	Certainty Case	Lottery Case	Reading	Math	Writing
2 years	10	1	7.2	7	7
1 year	0	18	6.1	6.7	10.1
Never	90	81	86.7	86.3	82.9

Panel B: # of years	% of Schools in Bottom 25th percentile				
	Certainty Case	Lottery Case	Reading	Math	Writing
during 1998 & 1999	Certainty Case	Lottery Case	Reading	Math	Writing
2 years	25	6.3	20.4	19.1	18.5
1 year	0	37.5	10.1	13.3	20.1
Never	75	56.3	69.5	67.6	61.4

Table B.3: Effect of the “Voucher Shock” program on treatment status
Checking robustness using different samples

(WRCT % above and % below scores, 1989-1997)

	WRCT % above				WRCT % below			
	66-47 FE (1)	60-47 FE (2)	66 FE (3)	60 FE (4)	66-47 FE (5)	60-47 FE (6)	66 FE (7)	60 FE (8)
Less treated * 1 year after program	3.28 (3.19)	2.55 (3.20)			0.94 (2.06)	1.53 (2.45)		
Less treated * 2 years after program	3.51 (2.67)	2.81 (2.67)			-0.79 (1.85)	-1.31 (2.16)		
Less treated * 3 years after program	4.25 (3.74)	3.24 (3.75)			0.86 (2.52)	-0.09 (2.98)		
Somewhat treated * 1 year after program	2.03 (2.81)	0.90 (3.02)	2.64 (2.63)	1.56 (2.75)	-0.54 (2.05)	1.29 (2.24)	-0.26 (1.94)	1.09 (2.04)
Somewhat treated * 2 years after program	5.38** (2.43)	4.57** (2.64)	5.22** (2.25)	4.24** (2.36)	-4.45* (1.88)	-4.28** (2.08)	-4.82** (1.78)	-3.66** (1.87)
Somewhat treated * 3 years after program	5.01 (3.03)	6.00** (3.21)	6.42** (2.90)	5.47** (3.01)	-2.60 (2.30)	-4.18*+ (2.49)	-3.99* (2.36)	-3.39 (2.38)
More treated * 1 year after program	-0.92 (3.33)	-0.72 (2.93)	-0.90 (3.32)	0.76 (2.92)	1.55 (2.50)	0.52 (2.16)	2.35 (2.47)	0.47 (2.15)
More treated * 2 years after program	6.06** (3.14)	6.35** (2.68)	5.94* (3.12)	6.32** (2.68)	-4.16* (2.50)	-5.27** (2.13)	-4.14* (2.48)	-5.21** (2.12)
More treated * 3 years after program	5.69 (3.98)	5.52 (3.47)	5.73 (3.97)	5.60 (3.46)	0.38 (3.16)	-0.51 (3.16)	-0.47 (3.15)	-0.62 (2.84)
Observations	1195	1195	1195	1195	1195	1195	1195	1195
R-squared	0.58	0.58	0.57	0.57	0.53	0.52	0.52	0.52

*, **, ***: 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 ethnicity, sex, percentage of students eligible for free or reduced-price lunches and real per pupil expenditure.

Free-Reduced Lunch eligible percentage of Milwaukee schools, 1990

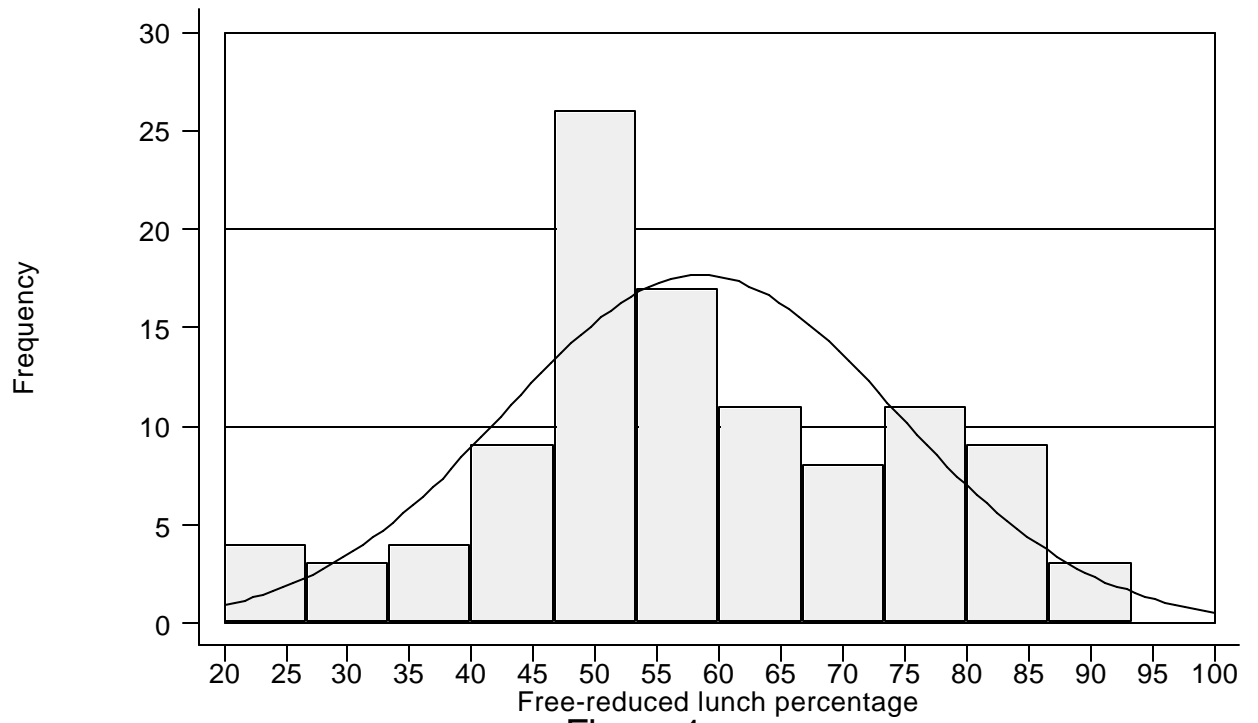


Figure 1

FCAT Reading, Math and Writing-F,D Schools

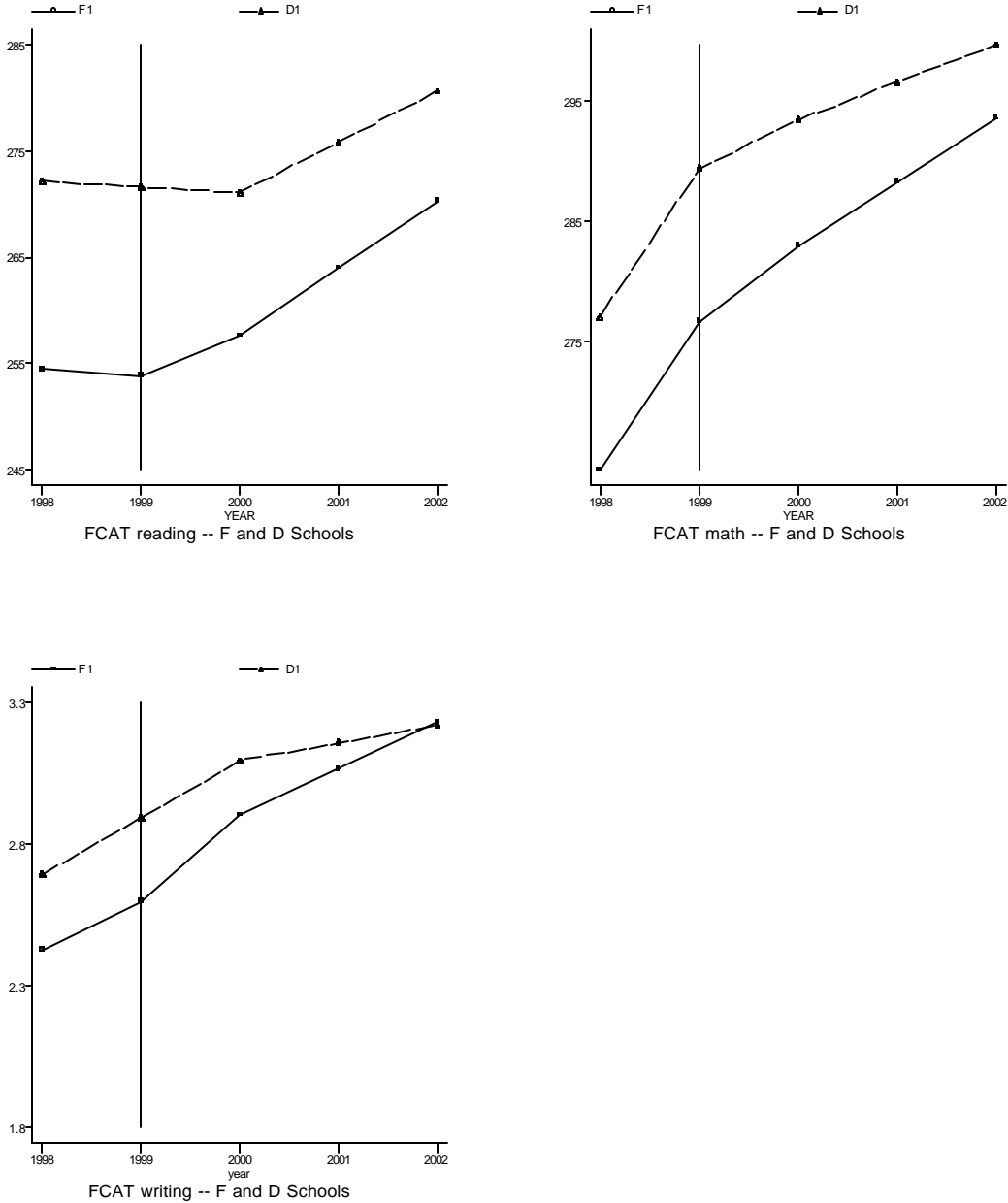


Figure 2. Florida 'threat of voucher' Program

ITBS Reading and Math --MT, ST and LT

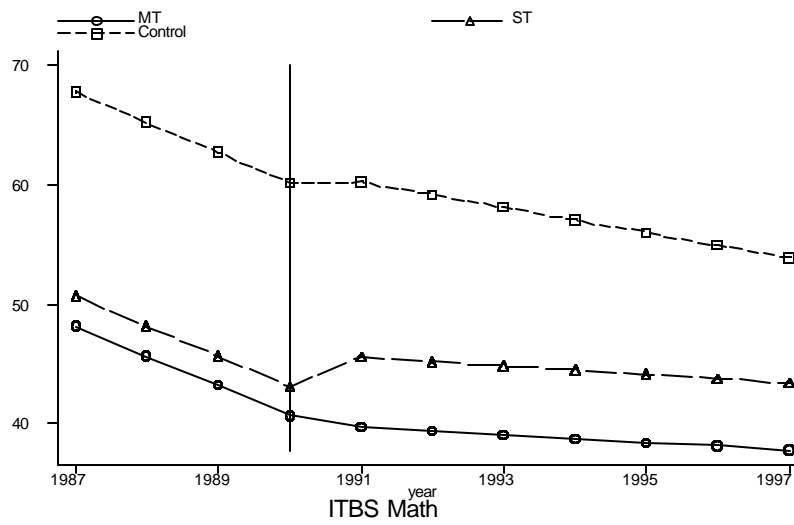
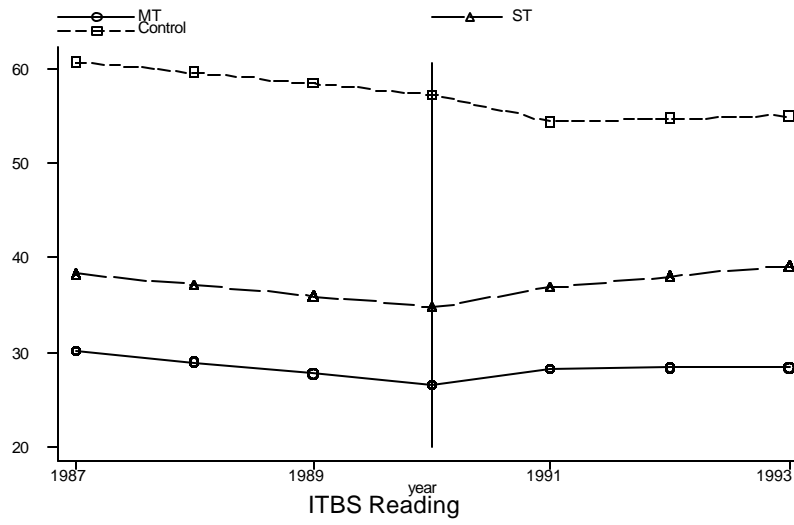


Figure 3. Milwaukee Voucher Shock Program